

-1199-

Query: 192 VRAHPIDELLKLYVRINAEAEFTDPVDEARENFREKLEANDPEATEIACQWFRDESILLEFN 251
V+AHPIDELLKLYVRINAEAEFTDP+VDEARENFREKLE D EATELAQWFRDESILLEFN
Sbjct: 181 VQAHPIDELLKLYVRINAEAEFTDPVDEARENFREKLEEDGKATELAQWFRDESILLEFN 240

5 Query: 252 RLYDQAVTFPDSYNGEAFYNDKMDVELELLESKNLLVESKGAQVVNLEKYGIEHPALIKK 311
RLYDQ++VTFPDSYNGEAFYNDKMDVE+LLE+KNLLVESKGAQVVNLEKYGIEHPALIKK
Sbjct: 241 RLYDQLHVTFPDSYNGSAFYNDKMDVEILLLEAKNLLVESKGAQVVNLEKYGIEHPALIKK 300

10 Query: 312 SDGATLYITRDLAALYRKRTYDFAKSIYVVGNEQSAHFQGLKAVLKEMGYDMSDMTHV 371
SDGATLYITRDLAALYRKRTYDFAKS+YVVGNEQ+AHFQGLKAVLKEM YDMSDMTHV
Sbjct: 301 SDGATLYITRDLAALYRKRTYDFAKSYVVGNEQAAHFQGLKAVLKEMGYDMSDMTHV 360

Query: 372 PFGLVTKGAKLSTRKGNVILLETVAEALNRAASQIEAKNPNLADKDKVAQVAVGVGAIK 431
PGLVTKGAKLSTRKGNVILLETVAEALNRAASQIEAKNPNLADK+ VA AVGVGAIK
Sbjct: 361 AFGLVTKGAKLSTRKGNVILLETVAEALNRAASQIEAKNPNLADKVAHVAHGVGAIK 420

15 Query: 432 FPDYKIDRTNGYDFDLEAMVSFEGTGPVQYAHARIQSILRKANFSPNSNDYSLDVE 491
FPDYLKTR NGYDFDLEAMVSFEGTGPVQYAHARIQSILRKA+FPS + YSL D E
Sbjct: 421 FPDYLKTRNGYDFDLEAMVSFEGTGPVQYAHARIQSILRKADPTPSATTTSLADAE 480

20 Query: 492 SWEIHKLIQDFPRIIVRAADNFEPISIAXFALNLAQCFMKYAYHTRILDEDAESRRLAL 551
SWEIHKLIQDFPRII R +DNTEPSI+AKFALNLAQ FHKYAYHTRILD+++E +RLAL
Sbjct: 481 SWEIHKLIQDFPRIIKRTSENTEPSIMAKFALNLAQSFHKYAYHTRILDNSERDRRLAL 540

25 Query: 552 CYATATVTLKESRLRLGVDAFNM 574
CYATATVTLKE+RLRLGVDAFNM
Sbjct: 541 CYATATVTLKALRLRLGVDAFNM 563

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1076

A DNA sequence (GBSx1150) was identified in *S. agalactiae* <SEQ ID 3315> which encodes the amino acid sequence <SEQ ID 3316>. This protein is predicted to be arginine hydroxamate resistance protein (argR). Analysis of this protein sequence reveals the following:

35 Possible site: 42
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.3252 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10269> which encodes amino acid sequence <SEQ ID 10270> was also identified.

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88596 GB:M18729 unknown protein [Streptococcus pneumoniae]
Identities = 63/141 (44%), Positives = 90/141 (63%)

50 Query: 4 MNKISRQKRIKRLIQSGQIQTBETICHLNKGIDVTQATLSRDLRLRIGLLKLSPEGL 63
M K + R + IK++I ++ TQ+EI+ L+ + VTQ TLSRDLRLRIGL K++ +
Sbjct: 1 MRGRDRHLQIKMTYTEERLSTQKEIQORLEAHNVCVTQTLSRDLRLRIGLTKVKKENMY 60

Query: 64 YVSLSTATSNRFSFALRSYILKVSASFMLVINTNLGRASVIANFIANDKEGLPEILLTWFAG 123
Y ++ L ++ V+RA F LVL+T LGRASVIAN +D ILGT+AG
55 Sbjct: 61 YVIVNETREKIDLVEPLSHLELGVARASFTLVLRHTKLGRASVIANIVDVNRKDEWILSTVAG 120

Query: 124 ADTLAVICQNEIDIAKVFPEKL 144
A+TLAVIC++ +AK+ E L
Sbjct: 121 ANTLAVICRDQHVAKLMEDRL 141

-1200-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3317> which encodes the amino acid sequence <SEQ ID 3318>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3176 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 101/145 (69%), Positives = 121/145 (82%)

Query: 4 MNKIERQKRIKRLIQSGQIGTQESIKHLKNEGIDVTQATLSRDLREIGLLKLRSPGKL 63
      MNK+ERQ+IKR+IQ+ IGTOE+IK HL+ BGI VTQATLSRDLREIGLLKLR +GKL
Sbjct: 1 MNKMERQQQIKRIIQEHIQTQEDIKNHLKGBGIVVTQATLSRDLRRIGLLKLRDQGGK 60

Query: 64 YYSLSSTNSRPSPALRSYILKVSRAFMVLVNTNLGEASVLNFIENDEKGLPELLOTMG 123
      YYSLS + PSP +R Y+LKV RA FMLVL+TNLGEA VLN ID + +ILGT+AG
Sbjct: 61 YYSLSPEVATPSPSEVRFYVLKVDRAFGMLVLRNMLGEADVLANLIDNDATBDILGTAG 120

Query: 124 ADTLNIVICNEDIKVFKEKLSVGL 148
      ADTLNVIC+E+IAK FEK+L+ GL
Sbjct: 121 ADTLNIVICREELAKRFKDLAAGL 145

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1077

A DNA sequence (GBSx1151) was identified in *S.galactiae* <SEQ ID 3319> which encodes the amino acid sequence <SEQ ID 3320>. This protein is predicted to be DNA mismatch repair protein hexa (mutS). Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3570 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA88597 GB:M18729 mismatch repair protein [Streptococcus pneumoniae]
Identities = 593/858 (69%), Positives = 698/858 (81%), Gaps = 14/858 (1%)

Query: 1 MAKPTISPGMQQYLDIKENYPDAFLFRMGDFYELFYDDAVKAAQILSILTSRNKNAEK 60
      MA +SPGMQQY+DIK+ YPDAPLLFRMGDFYELFY+DAV AAQILSILTSRNKNA+
Sbjct: 1 MAIEKLSPGMQQYVDIKKQYPDAPFLFRMGDFYELFYEDAVNAAQILSILTSRNKNADN 60

Query: 61 PIPMAGVFPYRSAQQYIDVLVVELGYKVAIARQMEDPKKAVGVVKEVVQVITPGTVVVESTK 120
      PIPMAGVFPYRSAQQYIDVL+E GYKVALASQMEDPK+AVGVVKEVVQV+TPGTVV+S+K
Sbjct: 61 PIPMAGVFPYRSAQQYIDVLIEQGYKVALASQMEDPKQAVGVVKEVVQVITPGTVVDSSK 120

Query: 121 PDSANNFLVAIDSDQDQTFGLAYMDVSTGEFQATLTFDSFVRSIELNLKAREIVVGYQL 180
      PDS NNFVL+ID + Q FGLAYMD+ TG+ F T L DF V EI NLKARE+V+GY L
Sbjct: 121 PDSQNNFLVSDIRBSNQ-FGLAYMDLVTDGTFVIGLDFITVCGRIENLKARKVVLGYDL 179

Query: 181 TDEKNELLTKQNNLLSYEDERLNDIHLDEGLTDLSEIAAEKLLQVHRTQKRELHLQ 240
      ++E+ +L++QNNL+LSYE E D+HL+D +L +E +A+ KLLQYVHRTQ REL+HL+

```

-1201-

Sbjct: 180 SEEEKILSRQNLVLSYEKESPEDLHLLDLRLATVETASSKLLQVHRTQMKELNHLK 239
 Query: 241 KVVHYEIKDYIQLMSVATKNSLDLLENARTSKKHGSLYMLDETKTAMGTSMILRTIDRPL 300
 V+ YEIKD+LQM YATK SIDL+ENAR+ KK GSI+MLDETKTAMG R+LR+MI RPL
 Sbjct: 240 PVIRYEIKDFLQMDYATKASLDLVENARSGKKQSSFLMLDETKTAMGSELLRSWIRPL 299
 Query: 301 VSMRIKERQDILIQVLDYFFERNDLTESLGVYDIERLASRVSPGKANPKDLQIGQTL 360
 + RI +RQ++QVFLD+FFER+DLT+SLKGVYDIERLASRVSPGK NPKDLQLQ TL
 Sbjct: 300 IDKERIVQRQEVVQVFLDHFERSDLTDSLGVYDIERLASRVSPGKINPKDLQLATTL 359
 Query: 361 SQIPRIKILQSGFNQPHLIDIVNKKDITMPELESINTAIAPAEQAQITWGNIIKSGFDQK 420
 S+PRI+ IL+ QP L ++ +D +PEESLI+ AIAPEA IT+G II++GPD+
 Sbjct: 360 SSVPRIRALEGMEQPTIAYLIAQDAIPELESLSAAIAPEAPHVITGXIIIRTPDET 419
 Query: 421 LDNYRTVMREGTMIADTEAKERNASGIGITLKIDYNNKDDGYYPHVINSNLSLVAPEHFRK 480
 LD YK V+REGT WIA+TEAKR SGI TLKIDYNNKDDGYYPHVINS L VP HFRK
 Sbjct: 420 LDYTRCVLRGTSWIAIEAKERENSGISTLKIDYNNKDDGYYPHVINSQLGNVFAHFRK 479
 Query: 481 ATLKNSERYGTAEALRIGBEMLEAREGSSNLEYDIPMRVRAQVESYIKRLQELAKTIATV 540
 ATLKNSE+GT ELA+IBG+MLEARE+S+NLEY+IPMR+R +V YI+RLQ LA+ IATV
 Sbjct: 480 ATLKNSEFGTSEARIBGDMLEAREKSANLEYIPMRIREEVGKVIQRQLAQGLATV 539
 Query: 541 DVLGSLAVVAENYHYVFRPKNDQHQIKIKNGRHATVEKMGVQVEYIPNSIYFDSQDIDQL 600
 DVLGSLAVAE +R+P+P+D QI I+ GRHA VEKMG Q YIR+I T IQL
 Sbjct: 540 DVLGSLAVVAETQHLIRPEFGDSDQIDIRGRHAEVVEKMGQYTIYIRWITQMAEDTSIQL 599
 Query: 601 ITGPNMSGKSTYMRQLALTVMACMGGFVSADDEVLPVDAIFTRIGAADLISQGSTPM 660
 +TGPNMSGKSTYMRQLA+T +MAQ+G +V A+ LP+PDAIFTRIGAADL+SQGSTPM
 Sbjct: 600 VTGPNMSGKSTYMRQLAMTAMQQLSGSYVPAESAHLIPDAIFTRIGAADLVSGQSTPM 659
 Query: 661 VEMMEANQAVKRASDESLILFDLGRGTATYDGMALQSIIEYIHORVAKTNIPATHYHE 720
 VEMMEAN A+ A+ SLILFDLGRGTATYDGMALQSIIEYIH+ +AKT+PATHYHE
 Sbjct: 660 VEMMEANNAISHATNSLILFDLGRGTATYDGMALQSIIEYIHEHIGAKTLIPATHYHE 719
 Query: 721 LITLSEBQTLRIYNVHVATLERDGEVYFLHKIESGPADKSYGIHVAKIAGLPDILLDRATD 780
 LT L L LYNHVATLE+DG+VFLPINKIE GPADKSYGIHVAKIAGLP DIL RA
 Sbjct: 720 LITLSESLQHLRIYNVHVATLEQDQVYFLHKIEGPADKSYGIHVAKIAGLPADLJARADK 779
 Query: 781 ILTQLEADAVQLIVSPSGEAVTADINEELDSEKQGGQLSPPEPSNAGRVIELEADIM 840
 IL+QLR + SP T+ + E Q+SLF+ + ++ EL +D+
 Sbjct: 780 ILTQLENQOTE---SPPPMRQTSATVE-----QISLFDR-AEESHPLAELAKLDVY 826
 Query: 841 NUTPMQVMNAIFDLKLL 858
 N+TPMQ MN + +LK+ L
 Sbjct: 827 NMTPMQVMNVLELKQNL 844

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3321> which encodes the amino acid sequence <SEQ ID 3322>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 532 - 548 (532 - 549)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 661/858 (77%), Positives = 746/858 (86%), Gaps = 7/858 (0%)
 Query: 1 MAKPTISPGMQYLDIKENYDPAFLFRMGDFYELPYDDAVKAAQLRISLTSRNQNAEK 60
 MAK ISPGMQYLDIK+YDPAFLFRMGDFYELFY+DAVKAQA+LRI LTSRNQNAE
 Sbjct: 1 MAKINISPGMQYLDIKKUYDPAFLFRMGDFYELFYEDAVKAAQLRIGLTSRNQNAE 60
 Query: 61 PIPMAGVFYHSAQYQYIDVLVELGYKVAIEQMEDPKCAVGVVRRVGVVTFGTGVVESTK 120

-1202-

		PIPMAGVP+HSAQQYIDVL+ELGYKVA+ARQMEDPK+AVGVVKREVQV+TPCTIV+S K
	Sbjct:	61 PIPMAGVPHSAQQYIDVLELGYKVAARQMEDPKQAVGVVKREVQVITPKTVVDSAK 120
5	Query:	121 PDSANFLVAIDSDQQCTGLAYMOVSTGEFQATLLTDPESVRSILNAKAREIVVQYL 180
		PDSANFLVA+D D +GLAYMOVSTGEF T L DP SVRSI NLAKA++++G L
	Sbjct:	121 PDSANFLVAVIDF-DGCRYGLAYMOVSTGEFCTDLADFTSVRSIQLNAKAREVLLGFDL 179
10	Query:	181 TDERKHLLTKQNNLLSYRDERNDTHLIDQLTDLTLEISAARKLAQYVHRTQKRELSHQ 240
		++H+ +L KQNNLLSYR+ D LID QLT +H+AA KLAQYVH+TO RELSHQ
	Sbjct:	180 SEEBQTLIVKQNNLLSYRSTVYEDKSLIDQLTIVELTANGKLAQYVHRTQKRELSHQ 239
	Query:	241 KVHYBIKDYLOMSYATKNSLDLLEWARTSKHGSLYMLDETAKTMRLEMTWIDRPL 300
		+VHYBIKDYLOMSYATK+SLDL+EWART+KKGSLYMLDETAKTMRG R+LR+WIDRPL
	Sbjct:	240 ALVHYBIKDYLOMSYATKSSLDLVEKARTNKGSGLYMLDETAKTMRMLLEWIDRPL 299
15	Query:	301 VSNRIKERQDIIQVFLDYFFERNOLTESLKGVDIERLASRVSPGKANPKDLLQGLT 360
		VS I ERQ+IIQVFL+ F ER DL+ SLKGVDIERL+SRVSPGKANPKDLLQGLT TL
	Sbjct:	300 VSKAILERQEIIVFLNAPIERTDLNSLKGVDIERLSRVSPGKANPKDLLQGLHT 359
20	Query:	361 SQPRIMILQSPNOBELDIIVNKDITMPELESINTAIAPRAQAQNTBGNIKSGFDRLQ 420
		+Q+P IK IL+SP+ P +D +VN ID++PELE LI TAI P+A ATI+EG+II+GPD++
	Sbjct:	360 AQVPYIKAILSEFSDPCVDKLVNDISLPELEYLIRTAIDPADAPATIBSISIRNGFDER 419
25	Query:	421 LDNYRTVMREGTGWIADIKAERAAAGIGTLKIDYNKDGYYPHVNSNLSLVEHFFRK 480
		LD+YR VMREGTGWIADIKAER ASGI LKIDYNKDGYYPHVNSNLSLVEHFFRK
	Sbjct:	420 LDHYRVMREGTGWIADIKAERQASGINLKDYNKDGYYPHVNSNLSLVEHFFRK 479
	Query:	481 ATLANSERYGTAEIAKIBEGMLEAREBQSSNLEYDIPMRVPAQVESYIKRLQELAKTIATV 540
		ATLANSERYGTAEIAKIBEG+MLEARE+SS+LEYDIPM +RAQVE+YI RLQ+LAK +ATV
	Sbjct:	480 ATLANSERYGTAEIAKIBEGMLEARESSSLEYDIPMCIRPAQVETYINRLQELAKIATV 539
	Query:	541 DVLQSLAVVAENHYHVRKPNDDHQIKIKNGRHATVEKVMGQVEYIPNSIFDQQTSLQ 600
		DVLQSLAVVAE HY+RP+PND H I I+ GRHA VEKVMGQVEYIPNSI FU QT IQL
	Sbjct:	540 DVLQSLAVVAETNHYRPFQNDNHNVTIQSGRHAVEKVMGQVEYIPNSISFDQQTSLQ 599
35	Query:	601 ITGPNMSGKSTYMRQLALTIVMAQMGFVSADVDLPVFDIAIFTRIGAADDLISGQSTFM 660
		ITGPNMSGKSTYMRQLALTIVMAQMG FV+AD VDLP+FDIAIFTRIGAADDLISGQSTFM
	Sbjct:	600 ITGPNMSGKSTYMRQLALTIVMAQMGSFVAADHVDLPFDIAIFTRIGAADDLISGQSTFM 659
40	Query:	661 VEMMEANQAVIKRASDKSLILFDELGRGTATYDGMALAQSIIEYIHDRVAKTMFATHYHE 720
		VEMMEANQA+KRASD SLILFDELGRGTATYDGMALAQ+IIEYIHDRV AKT+PATHYHE
	Sbjct:	660 VEMMEANQAIKRASDNSLILFDELGRGTATYDGMALAQAIIEYIHDRVAKTIFATHYHE 719
45	Query:	721 LTDLSBLTRLNVNVALEKRGDEVTLPHKIESGPAKDSYGIHVAKIAGLEPDLDRATD 780
		LTDLS L TVNVNVALE+DG+VTLPHKI GRADKSYGIHVAKIAGLE LL RA +
	Sbjct:	720 LTDLSNLTSLNVNVALEKRGDEVTLPHKIAEGPAKDSYGIHVAKIAGLEPDLKRADE 779
	Query:	781 ILSQLAEADAVQLIVSPSGRAVDADLNBEELDEKQOGLSLFPEPNSAGRVIELEPAIDM 840
		+L++LE S S E ++ E S +QOGLSLF + A + + L E ID+M
	Sbjct:	780 VILTRLETQ-----SRSTIISIVSPSQVSSSAVRQOGLSLFPEPNSAGRVIELEPAIDM 833
	Query:	841 NLTPMQNMNAIFDLKKLL 858
		N+TP+QAM +++LKKLL
	Sbjct:	834 NMTPLQMTTLYELKKLL 851

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1078

A DNA sequence (GBSx1152) was identified in *S. agalactiae* <SEQ ID 3323> which encodes the amino acid sequence <SEQ ID 3324>. This protein is predicted to be cold shock protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

-1203-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2095 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA869404 GB:A91080 unnamed protein product [unidentified]
 Identities = 48/63 (76%), Positives = 56/63 (88%)

Query: 1 MTQGTWKWPNSEKGGPGFISSETGTDVFAHPSSEIKVDGFKTLEBQKVTFFDIQDGGQRGPQA 60
 MT+GTWKWPN+KGGPGFI+SE G DVFAHPS+I+ GFKTL+EBQKVTFFD++ GQRGPQA
 Sbjct: 1 MTKGTWKWPNPDKGGPFITSBQGGDVFAHPSQIQTSGFKTLEBQKVTFFDVAGQRGPQA 60

Query: 61 TMI 63
 NI
 Sbjct: 61 VMI 63

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3325> which encodes the amino acid sequence <SEQ ID 3326>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2350 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 49/63 (77%), Positives = 56/63 (88%)

Query: 1 MTQGTWKWPNSEKGGPGFISSETGTDVFAHPSSEIKVDGFKTLEBQKVTFFDIQDGGQRGPQA 60
 M QGTWKWPN+EKGGPGFIS+E G DVFAHPS I+ +GFKTLEBQKVT FF+++GQRGPQA
 Sbjct: 3 MAQGTWKWPNSEKGGPGFISTENGQDVFAHPSAIQTNGFKTLEBQKVAFFDVESGQRGPQA 62

Query: 61 TMI 63
 NI
 Sbjct: 63 VMI 65

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1079

A DNA sequence (GBSx1153) was identified in *S.agalactiae* <SEQ ID 3327> which encodes the amino acid sequence <SEQ ID 3328>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.6378 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1204-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1080

A DNA sequence (GBSx1154) was identified in *S. agalactiae* <SEQ ID 3329> which encodes the amino acid sequence <SEQ ID 3330>. This protein is predicted to be DNA mismatch repair protein hexb (mutL).

Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2242 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10267> which encodes amino acid sequence <SEQ ID 10268> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AA86600 GB:M29686 mismatch repair protein [Streptococcus pneumoniae]
Identities = 452/657 (68%), Positives = 543/657 (81%), Gaps = 8/657 (1%)

Query: 20  LSKIIELPDLILANQIAGEVVRFPSSVVKELVENAIDAGSSQITIEVSEGLKIKQITDN 79
+S IIRLP++LANQIAGEV+ERP+SV KELVENAIDAGSSQI IE+EE+GLKK+QITIN
Sbjct: 1  MSHIIEPMEILANQIAGEVLRFPASVCKELVENAIDAGSSQIIIRIEAGLKKVQITIN 60

Query: 80  GBSMTSEDAVLSLRHATSKIKSQSDFRITRLGFRGEALPSIASISLMTIKTATECGKQ 139
G G+ ++ L-LRRHATSKIK+Q+DFRITRLGFRGEALPSIAS+S++T+ TA +
Sbjct: 61  GHGIAHDEVELARRHATSKIKNQADLFRITRLGFRGEALPSIASVSVLTLTAVDGASH 120

Query: 140  GTLLVAKGGNIEKQEVSSPROTKILVENLFFMTPARLKYMSQLSELAHIIDIVNRSL 199
GT LVA+GG +E+ +SP QTK+ VE+LFFMTPARLKYMS Q+EL+HIIDIVNR L
Sbjct: 121  GTKLVARGGVEEVI PATSFVQTKVCVEDELFTNTPARLKYMSQQARLSHIIDIVNRGL 180

Query: 200  AHPEVAFTLINDGKEMTKTSOTGDLRQAIAIGIATAKKMIISNADLDPEISGVVSLP 259
AHPE++F+LI+DGKEMT+T+OTG LRQAIAIGIQL +AKKMI EI N+DLDFEISG+VSLP
Sbjct: 181  AHPEFISLSIDGKEMTKTAGTQGLQAIAIGIYGLVSAKKMIETESDLDPEISGFVSL 240

Query: 260  ELTRANRNYITLLINGRYIKNPLNLSILDGYSGLIAVORFPPIAVIDIQIDPYADVAHV 319
ELTRANRNYI+L INRYIKNPLNR+ILDG+SGKLVORFP+AVI I IDPYADVAHV
Sbjct: 241  ELTRANRNYISLFNGRYIKNPLNRALIDGFGSGLIAVORFLAVHIIHIDPYADVAHV 300

Query: 320  PTKQEVRIKRELAELSLTAISLSLKQYDLIDALENLAKTSTVSVDKPIQTSPSLKQP 379
PTKQEVRIKRE+ELM+L+S AI+ SLK+ LIPDALENLAK++ R+ +K QT LK+
Sbjct: 301  PTKQEVRIKREKMLTIVLSKAIANSLEKQTLIPDALENIAKSTVVRNRKVBQTLIPLEN 360

Query: 380  GLYVDRAKNDFPIGAUTVSPIANFTNLKSDGSDVDVDMVSVAQCATQSPNIKYASRDQ 439
LIY++ + + +E L + K +++ T+ + +A R
Sbjct: 361  TLYYEKTEP----SRPSQTEVADYQVELITGQQDLTFPAKETLDR-LTKPAKLHFAEKRP 415

Query: 440  ADSENFHSQDYLSSQSLNKLVEKIDSEKSTTFPELEFFQMGQTYTLFAQGNGLYIID 499
A+ + E + L+ S+K +KL+ EE+S+PPELEFFQMGQTYTLFAQ GLYIID
Sbjct: 416  ANYDQLDHPELDIA---SIDKAYDKLEREEAGSPPELEFFQMGQTYTLFAQGRDGLYIID 472

Query: 500  QHAQGERVKYKYRREKIGEVNLSLQQLVPELFFPSSDFLQQRKMSLLQDWGVIPEFY 559
QHAQGERVKCYE YRE IG VD S QQLVPE++PEF + D L+L+R+M LL++W+EL Y
Sbjct: 473  QHAQGERVKYEEYRESIGNVDSQQQLVFPYIFEPADADALRKERNPLLEWGVFLAEY 532

Query: 560  GNTVFLRHPINWKEEVRSGIYENCMILLTINRVSVKYKRAELIWMCKRSITKANHT 619
G N FILRHPINW EEB+BSGIYENCMILLLT EVS+KCYKRAELIWMCKRSITKANH
Sbjct: 533  GENQFILRHPINWAEETESGIYENCMILLTKEVSIKCYKRAELIWMCKRSITKANHR 592
```

-1205-

Query: 620 LDYDSARHLLDQLAOCKNPYNCPHGKRPVLVNFTHKADMEKMFRIQENHTSLRLDLGKY 676
 +DD+SR LL QL+QC NPYNCPHGKRPVLV+FTK+DMKMF+RIQENHTSLR+LGKY
 Sbjct: 593 IDHDSARQLYLQLAQCKNPYNCPHGKRPVLVHFTESDMKMFRIQENHTSLRLDLGKY 649

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3331> which encodes the amino acid sequence <SEQ ID 3332>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1854 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 502/663 (75%), Positives = 574/663 (85%), Gaps = 9/663 (1%)

- Query: 20 LSKIIELPDILANQIAAGEVVERPSSVVKELVENAIDAGSSQITIVVEESGLKIKQITDN 79
 ++ IIELP+LANQIAAGEVVERP+SVVKELVENAIDAGSSQIT+*+RESGLK IQ+TDN
 Sbjct: 14 MTNIIELPEVLANQIAAGEVVERPASVVKELVENAIDAKSSQITVIEESGLKIMIQITDN 73
- Query: 80 GEGMTSEDAVLSLRHATSKIKSQSLFRTITLGFGRGALPSIASISIMTIKTATQCKQ 139
 GEGM+ ED LSLRRHATSKIKSQSLFRTITLGFGRGALPS+ASIS +TIKTAT+
 Sbjct: 74 GEGMSHEDPLSLRRHATSKIKSQSLFRTITLGFGRGALPSVASISKITIKTATKEVTH 133
- Query: 140 GTLLVAKGNGNIEKQEVVSPRGTKILVENLFFNTPARLKYKMSLQSELAHIIDIVNRLSL 199
 G+L+A GG IE E +S+P GTKI VENLF+NTPARLKYKMSLQ+ELAHI+D+VNRSL
 Sbjct: 134 GSLLIATGGEIETLEAISTPTGTTLKIVNLFNTPARLKYKMSLQAELAHIVDVNRSL 193
- Query: 200 AHPEVAFITLNDGKEMTKTSGTGDLRQAIGIYGLNTAKKMITISNADLFEISGVVSLP 259
 AHPEVAFITL+DG+++T+TSGTGDLRQAIGIYGLNT KKM+ ISNADLFE+SGVVSLP
 Sbjct: 194 AHPEVAFITLSDGRQLATSGTGDLRQAIGIYGLNTTKMLAISNADLFEVSGVVSLP 253
- Query: 260 ELTRANRNYITLLINGRYIKFNFLNRSILDGYSKLMVGRFPPIAVIDIQIDPYLADVNVH 319
 ELTRANRNY+T+L+NGRYIKFNFLNR+ILDGYSKLMVGRFPPI VIDIQIDPYLADVNVH
 Sbjct: 254 ELTRANRNYMTILVNGRYIKFNFLNRAILDGYSKLMVGRFPPIVIDIQIDPYLADVNVH 313
- Query: 320 PTKQEVRIKSKRELMLISTAISLSLKYDILPDALENAKTSTRSDKPIQTSFSLQEP 379
 PTKQEVRIKSKRELML+LISTAISLSL+ DLIPDALENLAK+STR KP QT L+
 Sbjct: 314 PTKQEVRIKSKRELMLISTAISLSLKEQDILPDALENLAKSSTRFSKPEQITQLPQSR 373
- Query: 380 GLYIDRAKNDFIADTVSEPIANFNIDKSDGVDNDVKNV-----NQGATQSPNIK 433
 GLYID KNDFF+ VSE I D G+VDN VK ++ ++K
 Sbjct: 374 GLYIDPQRNDFFVKSASVSEKI---PRDIFYSGAVDINSVKVEKVLPHSSSEVIGPSSGVK 430
- Query: 434 YASRDQADSENFIHSQDYLSKQSIKMLVKIDSEESSTFPELEFPQMGHTYLFQAQNG 493
 +ACR Q H L ++Q L+++ +L++E S SFEEL+PQMGHTYLFQAQNG
 Sbjct: 431 HAGRPQNTFTTIDHNLDLKNNRQKLSQMLTRLENRGSQSVFELDYFGMGHTYLFQAQNG 490
- Query: 494 GLYIIDHQAQSRVKYRYREKIGEVNLSLQQLVPLFEFSSSDFLQERMSLLQDVG 553
 GL+IIDHQAQSRVKYRYR+KIGEVN+SLQQLVPL+LFEFS SDF+ LQERM+LL +VG
 Sbjct: 491 GLFIIDHQAQSRVKYRYRDRKIGEVNLSLQQLVPLFYLFSSGSDFINLQERMLLNISVG 550
- Query: 554 IFLEPYGNNTFILRHPHIMWKREESVSIYKMDMLLNEVSVKRYRARIAMNSCKRS 613
 IFLE YG+NTFILRHPHIMWKREE+ SG+YKMDMLLNEVS+K YRARIAMNSCKRS
 Sbjct: 551 IFLEVYGNNTFILRHPHIMWKREELASGYKMDMLLNEVSIETRYRARIAMNSCKRS 610
- Query: 614 IKNHETLDYDSARHLLDQLAOCKNPYNCPHGKRPVLVNFTHKADMEKMFRIQENHTSLRLDLGKY 676
 IKNH+LDYDSAR+LL QLAQC+NPYNCPHGKRPVL+NP+KADMEKMF+RIQENHTSLR+LGKY
 Sbjct: 611 IKNHSLDYDSARNLLQLAQQNPYNCPHGKRPVLINFSKADMEKMFRIQENHTSLRLDLGKY 673

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1206-

Example 1081

A DNA sequence (GBSx1155) was identified in *S.agalactiae* <SEQ ID 3333> which encodes the amino acid sequence <SEQ ID 3334>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3372 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1082

A DNA sequence (GBSx1156) was identified in *S.agalactiae* <SEQ ID 3335> which encodes the amino acid sequence <SEQ ID 3336>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -14.01 Transmembrane 176 - 192 ( 170 - 197)
INTEGRAL Likelihood = -8.07 Transmembrane 390 - 406 ( 387 - 412)
INTEGRAL Likelihood = -6.10 Transmembrane 271 - 287 ( 269 - 291)
INTEGRAL Likelihood = -6.00 Transmembrane 83 - 99 ( 82 - 101)
INTEGRAL Likelihood = -4.78 Transmembrane 51 - 67 ( 50 - 71)
INTEGRAL Likelihood = -2.92 Transmembrane 303 - 319 ( 302 - 320)
INTEGRAL Likelihood = -2.76 Transmembrane 363 - 379 ( 362 - 381)
INTEGRAL Likelihood = -2.39 Transmembrane 152 - 168 ( 151 - 169)
INTEGRAL Likelihood = -2.02 Transmembrane 325 - 341 ( 325 - 342)
INTEGRAL Likelihood = -1.65 Transmembrane 226 - 242 ( 226 - 242)
INTEGRAL Likelihood = -0.90 Transmembrane 24 - 40 ( 24 - 40)
INTEGRAL Likelihood = -0.27 Transmembrane 111 - 127 ( 111 - 127)

----- Final Results -----
bacterial membrane --- Certainty=0.6604 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10265> which encodes amino acid sequence <SEQ ID 10266> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA61918 GB:X89779 LmrP integral membrane protein [Lactococcus
lactis]
Identities = 145/401 (36%), Positives = 236/401 (58%), Gaps = 4/401 (0%)

Query: 9 VKKEFFALPKQLQRLRLRFISITVGSALFFPMANYVVOYFNLVITGLIIITQLSGFVAT 68
+KEF+ L K LQLR + F+ +F M +YY QY G+ +IGIL+ ++ ++ FVA
Sbjct: 1 MKKEFNLNKNLQRLGIVFLGAFSYGVTFSSMTIYYNQLGSAITGILLALSVAITVNG 60

Query: 69 LYGGHLSAMGRKKVVIIGSLLATIGWATITAAVNPVHITPHLTFVGIIILIEAHQFFYP 128
+ G +D GRK V++ G+++ +G A+ IA+N+P H+ P TF+ L+I +F
Sbjct: 61 ILAGEFADNRGRKPVVVFCTIIQLGALALASNLFGHVNWSTFIFNLISFGYINPIT 120

Query: 129 AYEAMTIDLINQNRFRVYFYGWLNIAWMI GSGIAGIFVDHFFRLILVLLIISAICC 188
A AM ID +N +NR+ V+ + YW N++V+LG+ + + F LL++LL+ +

```

-1207-

Sbjct: 121 AGNAMIIDASNAEHRKVVFMLDYWAQNLVSIIAGALGAWLFRPAPFALLVILLITVLVSP 180

Query: 189 FVVYFKFDET-KPQESTPKHKDGVLTGFYONYSQVLADKAFVYVTLGAIGSSVVMQLQVNY 247
F+ F ET KP T K D+ F+ Y VL DK ++++ I ++ + Q DN+

5 Sbjct: 181 FLTTFMHTSYKF---TVKVDKAEINI PQAYKTVLQDKTYMIFMGANLATTPIIQDFNF 237

Query: 248 FSVNLKQNFVVSILGHTTIGAKMLSLAVPTNLLIVLMTINKPTIENWPLKQLQLGS 307
V+L +F+ ++ G I G +M++ + +L+VLMTT+N+ ++W ++ I GS

10 Sbjct: 238 LPVHLSNSFKTITFMGFKYQGRLTYLIIACVLAVLMTTLNRLTKDMSHQGFING 297

Query: 308 LIQSGFGLNFIENLTHGAILIAMTFTFGEMINYPASQVLRAEMVSEKIGSYSGPLA 367
L GW+P+ TF I ZA +T GE+Y P+ Q L A+M KIGSYG AI

10 Sbjct: 298 LFMAIGMTFSLTITFTTPIPIAGIVYTLGEIVVTPSVQTLGADLMNPKIGSYNGVAAIK 357

15 Query: 368 QPVASVLGAMVSLSYFTGKIGVQITLTITPLAGLVLINA 408
P+AS+LAG +VS+S IGW + L + + +L+L A

Sbjct: 358 MPTASILAGLLVSISSPMKKAIGVSLVIALTEVLAILIIVNA 398

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3337> which encodes the amino acid
20 sequence <SEQ ID 3338>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.41	Transmembrane	166 - 182 (161 - 188)
INTEGRAL	Likelihood = -7.75	Transmembrane	384 - 400 (376 - 403)
INTEGRAL	Likelihood = -7.64	Transmembrane	266 - 282 (261 - 285)
INTEGRAL	Likelihood = -4.25	Transmembrane	295 - 311 (291 - 313)
INTEGRAL	Likelihood = -2.71	Transmembrane	98 - 114 (98 - 115)
INTEGRAL	Likelihood = -2.23	Transmembrane	355 - 371 (355 - 374)
INTEGRAL	Likelihood = -2.02	Transmembrane	218 - 234 (218 - 234)
INTEGRAL	Likelihood = -1.91	Transmembrane	315 - 331 (315 - 331)
INTEGRAL	Likelihood = -1.22	Transmembrane	75 - 91 (75 - 92)
INTEGRAL	Likelihood = -0.75	Transmembrane	45 - 61 (45 - 63)
INTEGRAL	Likelihood = -0.75	Transmembrane	144 - 160 (144 - 161)

35 ----- Final Results -----

bacterial membrane ---	Certainty=0.5564 (Affirmative) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm ---	Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:CAA61918 GB:X89779 LmrP integral membrane protein [Lactococcus
lactis]

Identities = 138/400 (34%), Positives = 223/400 (55%), Gaps = 2/400 (0%)

45 Query: 1 MQEFLNLPKQIOLRLQVRFVTITLGSIFPMQMYTTFYGTFTGLAMITSLMGFVGT 60
M+EF NL K +QLR + F+ ++F M +YY Y G+ TG+L+ ++++ FV

Sbjct: 1 MKEPWMLDKQLRLQVIFLQPSYGVTPSSMTTYNQYLSAITGILLALGAVTFVAG 60

50 Query: 61 LYCHLSIDALGRKKVIMIGSVTTIGWLTITLANLPNAIPMLTFAGILVRIASSFGKP 120
+ G +D GRK V++ G++ LG L I +NLP PW TF L+ + +F

Sbjct: 61 ILAGFPADRNGRKPVMVFTIIQLLGAALALASNLPGHVNFGSTFIAPILISFGYNFVIT 120

55 Query: 121 AYEAWLIDLITDESRRFVYTYNYWFTINIAVMGRGLSGLPVDHIFLALLVALLVPLVCF 180
A AW+ID ++ NR+ V+ +YW N+V+ GA L + F ALIV LLL ++ F

Sbjct: 121 AGNAMIIDASNAEHRKVVFMLDYWAQNLVSIIAGALGAWLFRPAPERLIVILLITVLVSP 180

Query: 181 GVAYYCFDETRPETHAPDHGKLLASFPNYRQVPHDRAPVLPTLGAIFSGSINWQENYV 240
+ + ET T D + PQ Y+ V D+ ++F I + I MQ DN+

60 Sbjct: 181 FLTTFMHTSYKFVTVKVDKAEINI--PQAYKTVLQDKTYMIFMGANLATTPIIQDFNF 238

Query: 241 PVILKLYPQRTAVLGFQVTSKMSLMVLTNLLIVLPTVPMVNLKTEKMLLPQLVPSL 300
PVIL F+ GF++ +ML++ ++ +L+VL MT +N+LT+ W + GSL

60 Sbjct: 239 PVHLSNSFKTITFMGFKYQGRLTYLIIACVLVLLMTTLNRLTKDMSHQGFING 298

65 Query: 301 LFTLCMLLSPTFTQFVAINLSVLLITFGEMINVSASQVLRAEMVSEKIGSYSGPLA 360

-1208-

+GM+ SF T F I+++ ++ T GE++ + Q L AD+M+ +IGSY G ++
 Sbjct: 299 FMAIGMIFSLTITFTPIPIAGIVYTLGRIYVTPSVQ/TLGADLMPEKIGSINGVAIKM 358
 Query: 361 PLGAILASLIVSVSHPTGPLGQCLFAVIALGLIYPTVVS 400
 P+ +LA LNV+S +GV +A+ +L I +V+
 Sbjct: 359 PIASILAGLLVSIEMKAIQVSELVLAITEVLAITLVLA 398

An alignment of the GAS and GBS proteins is shown below.

Identities = 228/406 (56%), Positives = 305/406 (74%)

10 Query: 9 VKFEFALPKQLQRLLELRFISITVGSALPFPMMYVQVQNIATGILITQLSGFVAT 68
 ++EF LERQ+QLR+L+RF++IT+GS+IFFPMYMY YFG TG+L+IT L GFV T
 Sbjct: 1 NGEFLALPKQIQRLQVRFVITTLGSSIFFPMYMYITTTTGTFTWGLLWNTSLAGFVGT 60

15 Query: 69 LFGHGLSDAMGRKKVVIIGSLPATIGWAITIANVPHITTHLTFVGLLIEIAHQFVF 128
 LYOGHLSDA+GRKKV++IGS+ T+GW +TI AN+PH F LTF GIL++EIA FI P
 Sbjct: 61 LYOGHLSDALGRKKVIMIGSVGTTLGNFLTILANLPAALPMLTFPGILIVEIASFTGP 120

20 Query: 129 AYEANTILDTNQRNRFVYTTIGWLVMIAVMLGSGIAGIYFDHFFPELLIVLLISAIQC 188
 AYEAM IDLT+E NRRFVYTI YW +NIAVM G+G+G+FYDHF LL+ LL+++ +C
 Sbjct: 121 AYEAMIDLTDSRRRFVYTIYWFILIAVMFGAGLSGLFDHFFLALLVALLVNVICP 180

25 Query: 189 FVYVFKFDETKPQSGTQFKHKGVLQTFKRYSCVLVDKAFVYVTLGAIGSSVVMQLDNYF 248
 V Y+ FDET+P+ F H KG+L +F+NY QV D+AFV++TLGAI S ++Q+NY
 Sbjct: 181 GVAYYCFDETRFETHAFDGHGKLLASQNYRCVPHDRAFLVLTGLAIFSGSINWQDNYV 240

30 Query: 249 SVMKQNFVSVSILGHTITRQMLSLAVFTNILLIVLMTINKFVBNPLKROLLSL 308
 V+LK P+ ++LG +T +KMLSL V TMLLIVL MT +NK EWL QL++GSL
 Sbjct: 241 PVHLKLYPQPTAVLQFQVTSKMSLMLVLTNILLIVLMTVKNLTKWKLRLPVGSL 300

35 Query: 309 ICGFMPLNISNLTPGAILIAMFTPTGEMIVPASPQVLRANVMVEKIGSYSGFLAIQ 368
 + GML ++ F AI +++ TPGEMI V ASQVLR+NM +IGSY+GF++AQ
 Sbjct: 301 LFTLGMELSPFTTQFYAINLSVLLTPGEMINVSASQVLRANMDHSIGSTTGFVSAQ 360

Query: 369 PVASVLGAMVSLVYTTKIGVQITLTFMLAGLVLLIVATPMKI 414
 P+ +LA +VS+S+FTG +GVQ + L G+ ++ RMK +
 Sbjct: 361 PLGAILASLIVSVSHPTGPLGQCLFAVIALGLIYPTVVSAMKKV 406

A related GBS gene <SEQ ID 8725> and protein <SEQ ID 8726> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 SRCFLG: 0
 MoG: Length of UR: 4
 Peak Value of UR: 1.73
 Net Charge of CR: 1
 MoG: Discrim Score: -4.26
 GVH: Signal Score (-7.5): -2.48
 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 Amino Acid Composition: calculated from 1
 50 ALOM program count: 12 value: -14.01 threshold: 0.0
 INTEGRAL Likelihood = -14.01 Transmembrane 168 - 184 (162 - 189)
 INTEGRAL Likelihood = -8.07 Transmembrane 382 - 398 (379 - 404)
 INTEGRAL Likelihood = -6.10 Transmembrane 263 - 279 (261 - 283)
 55 INTEGRAL Likelihood = -6.00 Transmembrane 75 - 91 (74 - 93)
 INTEGRAL Likelihood = -4.78 Transmembrane 43 - 59 (42 - 63)
 INTEGRAL Likelihood = -2.92 Transmembrane 295 - 311 (294 - 312)
 INTEGRAL Likelihood = -2.76 Transmembrane 355 - 371 (354 - 373)
 INTEGRAL Likelihood = -2.39 Transmembrane 144 - 160 (143 - 161)
 60 INTEGRAL Likelihood = -2.02 Transmembrane 317 - 333 (317 - 334)
 INTEGRAL Likelihood = -1.65 Transmembrane 218 - 234 (218 - 234)
 INTEGRAL Likelihood = -0.90 Transmembrane 16 - 32 (16 - 32)
 INTEGRAL Likelihood = -0.27 Transmembrane 103 - 119 (103 - 119)
 PERIPHERAL Likelihood = 9.44 239
 65 modified ALOM score: 3.30

-1209-

icm1 HYPTD: 7 CFP: 0.660

*** Reasoning Step: 3

5 ----- Final Results -----

```
bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

10 The protein has homology with the following sequences in the databases:

ORF01675(325 - 1530 of 1854)
EGAD|041017|42348(1 - 400 of 408) integral membrane protein (LmrP) (Lactococcus lactis)
GP|1052754|cemb|CAA61918.1|X|89779 LmrP integral membrane protein (Lactococcus lactis)
PIR|S58131|S58131 integral membrane protein LmrP - Lactococcus lactis
%Match = 21.7
%Identity = 36.2 %Similarity = 60.8
Matches = 145 Mismatches = 155 Conservative Sub.s = 99

20 243 273 303 333 363 393 423 453
LQKLIVKICLNESKIKIQASGI*ENIDNYLLGKKGEYKVFEPALPQQLQRLSLRFISTVGSAPFPFMMYYVQYFQNL
:|||:| | |||:| : :| | | | |:|
MKSFVNLDRNLQLRLGIVFLGAFSPSTGVFSSMTIYYNQYLGSA
10 20 30 40

25 483 513 543 573 603 633 663 693
VTGILIIITQLSGFVNTLYGGHSLDAMCRKKVIIIGSLATTIGNAITTAANVPNHITPHITPVGLIIBIAHFQFPYAYE
| || | : : : | : : | : | : | : : : | : : | : : | : : | : : | : : | : : | : :
ITGIALLSAAVFAGMILAGLGFADNRGRKVVMTFTIQLGAALATASNLPGHVNPWSTPIAFLISPLFGYNIVTATGN

60 70 80 90 100 110 120

[illegible][illegible][illegible]

50
ASVLGAGMYSLSPTGKIQVITLIPMLAGILLYLATKMNIEIGK*NVRLY*KRIE*NGG*TYCCNWSWGIHIDICG
||:|||::||| ||| : : : ::||| |
ASILAGLLSLSPMKIKGVSLALTEVALTILLVAVNRHQTKLN
370 380 390 400

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1083

A DNA sequence (GBSx1157) was identified in *S.galactiae* <SEQ ID 3339> which encodes the amino acid sequence <SEQ ID 3340>. This protein is predicted to be holliday junction DNA helicase (ruvA).

60 Analysis of this protein sequence reveals the following:

Possible site: 37

-1210-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 75 - 91 (74 - 91)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:BA04943 GB:AP001511 holliday junction DNA helicase [Bacillus halodurans]
 Identities = 86/201 (42%), Positives = 122/201 (59%), Gaps = 6/201 (2%)

Query: 1 MYDYIKGKLSKITAKPIVETAGLGYMIYVANPYSFGYVNVQVITYLHQVIRDDAHLF 60
 M DY++G L I ++ VVE G+GY +Y NPY F + +TIY Q +R+D L+
 15 Sbjct: 1 MIDYLRGLTLDIDHQYAVVEHVGQVYVCNPFYFEKERDQSVITTYTFQVYREDVIRLY 60

Query: 61 GFHTENEKRIFLNLISVSGIGPTTALAIIVDDNGLVSAIDNSDIKYLTKFKIGKKT 120
 GF T+ ++ +F L++VSGIGE ALAI+A E ++ AI+ D +L KFP +GKKT
 20 Sbjct: 61 GFRTYKRSFLFEKLIVSGIGPKGALAILATGQPHVIALEEDAEAPLVKFPVGKKT 120

Query: 121 QQMILDLGKGFVE-----ASGESATSRKVSBNQNSLEAMEALLALGYKATLKKVKA 174
 +Q+ILD L GK E + E ++ N L+EAMEAL ALGY ELKVK
 Sbjct: 121 RQIILDLGKGVDELHGLFSQKEBQPKPHERNDGNQALDEAMEALKALGYVEKELKVKP 180

25 Query: 175 FFEGTNETVYQYIKSSLMML 195
 E T + YIK +L+++
 Sbjct: 181 KLSQETLTIDAYIKKALQML 201

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3341> which encodes the amino acid
 30 sequence <SEQ ID 3342>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.59 Transmembrane 75 - 91 (74 - 91)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 40 The protein has homology with the following sequences in the databases:

>GP:BA04943 GB:AP001511 holliday junction DNA helicase [Bacillus halodurans]
 Identities = 91/201 (45%), Positives = 128/201 (63%), Gaps = 5/201 (2%)

45 Query: 1 NYDYIKGKLSKITAKPIVETAGLGYMIYVANPYSFTDSVNLQVITYLHQVIREDAHLF 60
 M DY++G L I T +Y VVE G+GY + NPY F + +TIY Q +R+D L+
 Sbjct: 1 MIDYLRGLTLDIDHQYAVVEHVGQVYVCNPFYFEKERDQSVITTYTFQVYREDVIRLY 60

Query: 61 GFHTENEKRIFLNLISVSGIGPTTALAIIVDDNGLVSAIDNSDIKYLTKFKIGKKT 120
 GF T++++ +F KL++VSGIGE ALAI+A E ++ AI+ D +L KFP +GKKT
 50 Sbjct: 61 GFRTYKRSFLFEKLIVSGIGPKGALAILATGQPHVIALEEDAEAPLVKFPVGKKT 120

Query: 121 QQMILDLGKGFVE-----PQETGHTKARSNGKGTOLDEIKAILALGYKATLKKVKA 175
 +Q++LD L GK E Q+ K GN LDEA+EAL ALGY KELCK+
 55 Sbjct: 121 RQIILDLGKGVDELHGLFSQKEBQPKPHERNDGNQALDEAMEALKALGYVEKELKVKP 180

Query: 176 FFEGTSETABOYIKSALKML 196
 E + T + YIK AL+L++
 Sbjct: 181 KLSQETLTIDAYIKKALQML 201

- 60 An alignment of the GAS and GBS proteins is shown below.

Identities = 153/197 (77%), Positives = 176/197 (88%), Gaps = 1/197 (0%)

Query: 1 NYDYIKGKLSKITAKPIVETAGLGYMIYVANPYSFGYVNVQVITYLHQVIRDDAHLF 60

-1211-

```

MYDYIRK+L+KITAK+IVVTE GLGYMI VAMPYSP+ VMQ VTIYLIHQVIR+DAIILLF
Sbjct: 1 MYDYIRKGQLTKITAKYIVVEANGSLGYMINVANVSPFSDSVNQLVTIYLIHQVIREDAIILLF 60

Query: 61 GFHTTEKEKIFLNLI+SVSGIGPTTALAI+AVDNDNGLVSA+IDNSDIKYLTKPKPKGKCTA 120
5 GFHTTE+EK++FL LISVSGIGPTTALAI+AVDNDNGLV+AIIDNSDIKYL KFKPKGKCTA
Sbjct: 61 GFHTTEDEKQVFLKLI+SVSGIGPTTALAI+VAVDNDNGLVNA+IDNSDIKYLKMKFPKPKGKCTA 120

Query: 121 QQMLLDLQKGFVEASGESA-TSRKVSSEQSNLEHSMALLALGYKATELKKVKAPFET 179
10 QQM+LDL+GKPFVEA E+ T + + N+ L+EA+EALLALGYKA ELKK++APFET
Sbjct: 121 QQMVLIDLAGKGFVEAPQETQHTKARENKAGNTQLDEAIRALLALGYKAKELKKIRAPFET 180

Query: 180 NETVEQYIKSSLMIMK 196
+ET EQYIKS+LK+LMK
Sbjct: 181 SETAQYIKSALKLMK 197
15

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1084

A DNA sequence (GBSx1159) was identified in *S. agalactiae* <SEQ ID 3343> which encodes the amino acid sequence <SEQ ID 3344>. This protein is predicted to be DNA-3-methyladenine glycosidase I (tag). Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2812 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
25

```

A related GBS nucleic acid sequence <SEQ ID 10263> which encodes amino acid sequence <SEQ ID 10264> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AACT6573 GB:AE000432 3-methyl-adenine DNA glycosylase I,
constitutive [Escherichia coli K12]
35 Identities = 87/176 (49%), Positives = 122/176 (68%), Gaps = 1/176 (0%)

Query: 5 MKRCSWNLNDNPLYAYHDKKQWRAVHDDHVLPELLCLETYSGLSMETVIANKRQEPQV 64
M+RC WV+ D FLY+AYHD EWG D LFE++CLE Q+GLSW TVL KR+ +R
Sbjct: 1 MEROGWVQSD-PLYIAYHNDWGVPEYDTSKCLFPMICLEGQQAGLSWITVLKKRENTYRAC 59

Query: 65 PHNYTEKVAAMSDADLEILQNPRVIRHRLKLFSTRQARASIILOKEQSGFSFRYINWF 124
FH ++ KVAAM + D+E ++Q+ +IRHR K+ + NAR+ + +++ F ++WSF
Sbjct: 60 FHQFPDVKVAAMQBEDVERLVQDAGIRHRGKIQALIGNARAYLQMBQNGEPFVDVWVSF 119

Query: 125 VDNKQVQSVNNYNDVPASTLSERLSKDLKKRGKFKVGPPTCLYSFIQAAGMVNDH 180
V+++ QV +++P ST+ S+ LSK LKKRGKFPVG T YSF+QA G+VNDH
Sbjct: 120 VNHQFQVQATTLSEIPTSTGASDALSKALKKKRGKFKVGPPTTICYSFMQAQGLVNDH 175
45

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3345> which encodes the amino acid sequence <SEQ ID 3346>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4149 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
55

```

-1212-

An alignment of the GAS and GBS proteins is shown below.

Identities = 114/184 (61%), Positives = 135/184 (72%)

```

5  Query: 3  FHMGRCSWNLNPLVAVYHDKWGRAVHDDHVLPELLCLETYQSGLSWETVLNKRQSPR 62
    FHMGRCSWV DN LY YHD EMG+ + DD FELLCLLE+YQSGLSW TVL KRQ FR
    Sbjct: 2  FHMGRCSWPKDNQLYCYDHYLEWQQLDDDRDFPELLCLESYQSGLSWITVLAKKQAPR 61

10 Query: 63  QVFHHYNTKVAAMSDADLEIILQNRVIRHRKLPSTRQNRASILIQKEFGSFORYIY 122
    VFHHY-I VA + ++ LNP +IRH+LKL +T NA ++ TQKFGSF Y+W
    Sbjct: 62  TVFHHYDIASVATPTSSEMDALENPSIIRHKLKLATVNNALVQKIQKRFSGFSTYLM 121

    Query: 123 SFVNNKVCVNSVMNNYNDVPASTTILSERLSKDLAKRGFKFVGPTCLYSFQAAGMNDHEN 182
    +FV K N VN N VEA T LS RL+KDLAKRGFKF+GPT +YSP+QA+G+VNDHE
    Sbjct: 122 NFVGGKFINRLVQSENLVPAQTLSTIRLA+DLAKRGFKFLGPTTVYSFMAQAGLVNDHEE 181

15 Query: 183 ICDP 186
    C F
    Sbjct: 182 ACVF 185
  
```

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1085

A DNA sequence (GBSx1160) was identified in *S. agalactiae* <SEQ ID 3347> which encodes the amino acid sequence <SEQ ID 3348>. This protein is predicted to be competence-damage inducible protein (cinA). Analysis of this protein sequence reveals the following:

```

25 Possible site: 22
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
  
```

30 A related GBS nucleic acid sequence <SEQ ID 10261> which encodes amino acid sequence <SEQ ID 10262> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA64071 GB:Z34303 CinA protein [Streptococcus pneumoniae]
Identities = 194/297 (65%), Positives = 236/297 (79%), Gaps = 1/297 (0%)

40 Query: 1  MVEGSIPLQNLTLGLAVGGIVTSKGVQYMLGPPSELKPMWMEQVVPILSNNGTKLYSRV 60
    +VEG+IPL N TGLAVGG + GV Y+VLGPPSELKPMV+ Q++P L G+KLYSRV
    Sbjct: 121 IVEGAIPLPNETSLGLAVGGKLEVDGVTVVVLGPPSELKPMVLNQLLPKLMT-GSKLYSRV 179

    Query: 61  LRFPGIGESQLVTILEDIIRNQTDPTIAPYAKVGEVTLRLSTKAENQDEADFKLDSLEKE 120
    LRFPGIGESQLVTIL D+I NQ DPT+APYAK GEVTLRLSTKA +QEA+ LD LE +
    Sbjct: 180 LRFPGIGESQLVTILADLDINQIDPTIAPYAKTGEVTLRLSTKASSQREANQALDLENG 239

    Query: 121  ILALKTLNDRKLDLLGYGDDNNGMARTVLELLKVNKTIITAESITAGLFQSQLAFSG 180
    IL +T + L+D YGYG+ S+A V+E LK Q KTI AASSI/AGLFQ+ +A FSG
    Sbjct: 240 ILDCQTFEGISLRDFCYGYGEETSLASIVVEELRGKGTIAAASITAGLFQATVANFSG 299

    Query: 181  ASQVFNGSGFTTYSMEAKSQILGIPKKKLQSYGVVSHFTAEVMAQQARQLKADFGIGITG 240
    S +F GGF TYS+E KS++L IP K L+E+GVVS PTA+ MA+QAR ++DFGI LTG
    Sbjct: 300 VSSIFEGGFTTYSLEKSRMLDIPAKNLEHGVVSGEPTAQRMARQKRSQSDSGITG 359

55 Query: 241  VAGPDLEEGYPAGTVFPIGIATPEGVSSIKVSGKRSRDVRIHSTLHAFDLVRALL 297
    VAGPD LEG+P GTVPIG+A +G IKV+IGG+SR+DVRHI+ +HA+ +LVR+ALL
    Sbjct: 360 VAGPDSLEGHPTVPIGLAQDQGTETVLKVNIGRSRADVRHIAVMHAFNLVRKALL 416
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3349> which encodes the amino acid sequence <SEQ ID 3350>. Analysis of this protein sequence reveals the following:

Possible site: 22

```
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.91 Transmembrane 134 - 150 ( 134 - 150)

----- Final Results -----
bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CA84071 GB:Z34303 ClnA protein [Streptococcus pneumoniae]
Identities = 286/417 (68%), Positives = 336/417 (79%), Gaps = 1/417 (0%)

Query: 1 MKAEILAVGTEILTGGQIVNTNAQFLSEKMAKLGIDVVFQTAAGDNEERLLSVITASQRS 60
      MKAE+LAVGTEILTGGQIVNTNAQFLSEK+AE+G+DVVFQTAAGDNEERLLS++ ASQRS
Sbjct: 1 MKAEILAVGTEILTGGQIVNTNAQFLSEKLAEGIDVVFQTAAGDNEERLLSVITASQRS 60

Query: 61 NLVILCGGLGPTKDDLTQTLAKYLRKDLVYDEQAQCKLDDFFAKRKPSSRTFNNERQAQ 120
      +LVIL GGLG T+DDLTKQTLAK+L K LV+D QA +KLD PFA R +RTFNNERQAQ
Sbjct: 61 SLVILTGGLGATDDLTQTLAKFLKALVDFDPAQEKLDIFFALRPDYARTFNNERQAQ 120

Query: 121 VIEGSIPLFNKTGLAVGGFTITVDGISYVVLGPGPSELKPMVNEELVPLLSKQYSTLYSKV 180
      ++EG+IPLFN+TGLAVGG + VDG++YVVLGPGPSELKPMV +L+P L S LYS+V
Sbjct: 121 IVEGAIPFNKTGLAVGGKLVGVTYVVLGPGPSELKPMVNLQLLPKIMTG-SKLYSRV 179

Query: 181 LRFFGIGESQVLT+LSDPTIENQTDPTIAPYAKTGEVTLRLSTKTENQALADKKLGQLEAQ 240
      LRFFGIGESQVLT+L+D I+NQ DPT+APYAKTGEVTLRLSTK +Q A++ L LE Q
Sbjct: 180 LRFFGIGESQVLTILADLIINQIDPTIAPYAKTGEVTLRLSTKTASSQEEANQALDILENQ 239

Query: 241 LLSRKTLEGGQPLADVFGYGEINSLARETFELLVKYDKTITAESLITAGLPQSTLASFPG 300
      +L +T EG L D YGVGR+ SLA E L + KTI AAEISLITAGLPQ+T+A+P G
Sbjct: 240 ILDCQTFEGISLRDFCYGGEISLASIVVEELKRGKTIAAAEISLITAGLPQATVAFPSG 299

Query: 301 ASQVFNGGFVTYSMEEKAKMLGLPLEELKSHGVVSAYTAEGMAEQARLLTGADIGVSLTG 360
      S +F GGFVTYS+EEK+ML +P + L+ HGVVS +TA+ MABQAR T +D G+SLTG
Sbjct: 300 VSSIFEGGFVTYSLEEKSRMLDIPAKNLEHGVVSEPTAQMAEQARSKTQSDPGISLTG 359

Query: 361 VAGPDMLEBQAGTVFIGLATKQNKVESIKVLISGRSRDLVRYIATLHAFNMVRKTL 417
      VAGPD LE P GTVFIGLA E IKV I GSR DVR+IA +HAFN+VRK LL
Sbjct: 360 VAGPDSLEBHPVGTVFIGLAQQQGVTEVIKNIGSRADVRHIAVMHAFNLVRKALL 416
```

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/299 (67%), Positives = 242/299 (80%)

```
Query: 1 MVEGSIPLQNLGLAVGGIVTSKGVQYNVLGPGPSELKPMVMEQVVPILSNNGTKLYSRV 60
      ++EGSIPL N TGLAVGG +T G+ Y+VLGPGPSELKPMV E++VP+LS + LYS+V
Sbjct: 121 VIEGSIPLFNKTGLAVGGFTITVDGISYVVLGPGPSELKPMVNEELVPLLSKQYSTLYSKV 180

Query: 61 LRFFGIGESQVLTILEDIIKNQIDPTIAPYAKGEVTLRLSTKTAENQDEADFKLDEEKE 120
      LRFFGIGESQVLT+L D I+NQIDPTIAPYAK GEVTLRLSTK ENQ AD KL LE +
Sbjct: 181 LRFFGIGESQVLT+LSDPTIENQTDPTIAPYAKTGEVTLRLSTKTENQALADKKLGQLEAQ 240

Query: 121 ILAKTLINRKLKDLIAYGYNDSMARTVELLKVQNKTTITAESLITAGLPQSGLARFSG 180
      +L+ KTL+ + L D+ YGYG+NS+AR ELL +KTTITAESLITAGLPQS LA P G
Sbjct: 241 LLSRKTLEGGQPLADVFGYGEINSLARETFELLVKYDKTITAESLITAGLPQSTLASFPG 300

Query: 181 ASQVFNGGFVTYSMEAKSQLLGIPKKKLQYGVVSHPTAEAMQAQARQLLKADPGILGT 240
      ASQVFNGGF TYSME K++LIG+P ++L+ +GVVS +TAE MA+QAR L AD G+ LRG
Sbjct: 301 ASQVFNGGFVTYSMEEKAKMLGLPLEELKSHGVVSAYTAEGMAEQARLLTGADIGVSLTG 360
```

-1214-

Query: 241 VAGPDELRGYPAGTVFIGIATPEGVSSIKVSIIGKRSUVRHISTLHAFDVLVRALLKI 299
 VAGPD LE PAGTVFIG+AT V SIKV I G+SR DVR+L+TLHAF++VR+ ILK+
 Sbjct: 361 VAGPDMLREQBAGTVFIGIATQNKVESIKVLISGRSLDVRVYIATLHAFNVRVITLLKL 419

- 5 SEQ ID 3348 (GBS646) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 2-4; MW 61.6kDa), in Figure 134 (lane 3; MW 57.5kDa + lanes 2 & 4; MW 27kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 5-7; MW 36.6kDa) and in Figure 178 (lane 5; MW 37kDa).

GBS646-His was purified as shown in Figure 229, lane 5.

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1086

A DNA sequence (GBSx1161) was identified in *S.galactiae* <SEQ ID 3351> which encodes the amino acid sequence <SEQ ID 3352>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 59
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 148 - 164 (148 - 164)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3353> which encodes the amino acid sequence <SEQ ID 3354>. Analysis of this protein sequence reveals the following:

- 25 Possible site: 59
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 148 - 164 (148 - 164)
 ----- Final Results -----
 30 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 35 The protein has homology with the following sequences in the databases:

>GP:AD04860 GB:AF069745 RecA protein [Streptococcus parasanguinis]
 Identities = 333/381 (87%), Positives = 356/381 (93%), Gaps = 3/381 (0%)

- 40 Query: 1 LAKKLKQNEETTKQGERKALDNLKNIKEDFGKGAVMILGHRABQKQVMSSGSLAL 60
 +AKK KK ++ITKKGDER KAL+DALK IEKDPKG++MELGHRABQKQVMSSGSLAL
 Sbjct: 1 MAKKQKLLDDITKQGERKALDNLKNIKEDFGKGSIMRLGHRABQKQVMSSGSLAL 60
- 45 Query: 61 DIALGAGGYPKGRILIIETGPSSSKITTVLHAVAQAQKGGIAAFIDAHALDPAVAAL 120
 DIALGAGGYPKGRILIIETGPSSSKITTVLHAVAQAQKGGIAAFIDAHALDPAVAAL
 Sbjct: 61 DIALGAGGYPKGRILIIETGPSSSKITTVLHAVAQAQKGGIAAFIDAHALDPAVAAL 120
- 50 Query: 121 GVNIDELLISQPDSSGQGLIETAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDGSHVGLQ 180
 GVNIDELLISQPDSSGQGLIETAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDGSHVGLQ
 Sbjct: 121 GVNIDELLISQPDSSGQGLIETAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDGSHVGLQ 180
- 55 Query: 181 ARWASQAMRKLGASINKTKTTFIATFINDLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240
 ARWASQAMRK. ASINKTKTTFIATFINDLREKVGVMFGNPETTPGGRALKFYASVRLDVRG
 Sbjct: 181 ARWASQAMRKLGASINKTKTTFIATFINDLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240
- Query: 241 TTOIKGTGQDKDSSIGKETIKIVVKVAPPPKVAEVEIMYGBGISRTGELVKIASDLDI 300
 TQIKGTGQDKD++GKETIKIVVKVAPPPK A VEIMYGBGISRTGELVKIA+DLDI

-1215-

Sbjct: 241 NTQIKGTGDKQDITNVGKSTIKIVVKVKVAPPFKEMVEIMYSGISRTGELVKIATLDLI 300

Query: 301 IQKAGAWFSYNGEKIGQGSSENAKRYLADHPFLDEIDLKVRKVFGLLEESREESANA 360
 IQKAGAW+SYNGEKIGQGSSENAK++LADHPE+FDEID KVRV FGL+E+ E ++

5 Sbjct: 301 IQKAGAWFSYNGEKIGQGSSENAKRYLADHPFLDEIDLKVRKVFGLLEEKDEAVKSLDKTE 360

Query: 361 EE---TDDLALDLNGIIEID 378
 E +++ LDDL+ IEIED

10 Sbjct: 361 EAAPVVEEVTLDLDAIEIED 381

An alignment of the GAS and GBS proteins is shown below.

Identities = 339/379 (89%), Positives = 356/379 (93%), Gaps = 1/379 (0%)

15 Query: 1 MAKTKKARBITKKPGDERRKALDDALKNIEKDFGKGVNRLGERABQKVQVMSGSGLAL 60
 +AKK KK RRTTKPGDERRKALDDALKNIEKDFGKGVNRLGERABQKVQVMSGSGLAL
 Sbjct: 1 LAKKLKQNEBITKKPGDERRKALDDALKNIEKDFGKGVNRLGERABQKVQVMSGSGLAL 60

Query: 61 DIALGAGGYPKGRIVEIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL 120
 DIALGAGGYPKGRI+RIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL

20 Sbjct: 61 DIALGAGGYPKGRIEYIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL 120

Query: 121 GVNIDELLSSQDSGEGLEIAGKLIDSGAVDLVVDSVAALVPRAEIDGIDGSHVGLQ 180
 GVNIDELLSSQDSGEGLEIAGKLIDSGAVDLVVDSVAALVPRAEIDGIDGSHVGLQ

25 Sbjct: 121 GVNIDELLSSQDSGEGLEIAGKLIDSGAVDLVVDSVAALVPRAEIDGIDGSHVGLQ 180

Query: 181 ARMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYSSVRDLVRG 240
 ARMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFY+SVRDLVRG

30 Sbjct: 181 ARMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRDLVRG 240

Query: 241 NTQIKGTGDKHNVGKSTIKIVVKVKVAPPFRAEVEIMYSGISRTGELIKIASDLDI 300
 TQIKGTG+ KD ++GKETIKIVVKVKVAPPF+ A VEIMYSGISRTGEL+KIASDLDI

35 Sbjct: 241 TTQIKGTGDKQDSSIGKETIKIVVKVKVAPPFKVAEVEIMYSGISRTGELVKIASDLDI 300

Query: 301 IQKAGAWFSYNGEKIGQGSSENAKRYLADHPFLDEIDLKVRKVFGLLE-ESERESMANA 360
 IQKAGAW+SYNGEKIGQGSSENAK+YLAD+P +FDEID KVRV FG+ E +S +S +

Sbjct: 301 IQKAGAWFSYNGEKIGQGSSENAKRYLADHPFLDEIDLKVRKVFGLLE-ESERESMANA 360

Query: 361 KNEADLVDLDDNAIEIEE 379
 E DDL LDDLN IEIE+

40 Sbjct: 360 SEETDDLALDLNGIIEID 378

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1087

45 A DNA sequence (GBSx1162) was identified in *S. agalactiae* <SEQ ID 3355> which encodes the amino acid sequence <SEQ ID 3356>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2344 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 10259> which encodes amino acid sequence <SEQ ID 10260> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG37358 GB:AF028804 NrpR [Lactococcus lactis subsp. cremoris]
 Identities = 69/132 (52%), Positives = 102/132 (77%)

-1216-

Query: 5 MIKIYTISSCTSCCKAKTWLNAHQLPYKEQNLGKESLTRDEILEILTKTESGIESIVSSK 64
 MI IYT SCTSCCKAKTWL+ H +P+ E+NL + L+ EI +IL K +G+E ++SS+
 Sbjct: 1 MITIYTAPSCCTSCCKAKTWLSYHIFPNERNLADPLSTTETSQILQKCDGVEGLISSR 60

Query: 65 NRYAKALNCNIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLERSIRN 124
 NR+ K L + E++S+++ I +I ENP+I++ PI++D+KRL VGY E++IRAFLER++R
 Sbjct: 61 NRFVKTLGVDFEDISLSQAIIKENFQIMRRPIIMDEKRLHVGYNSEEIRAFLERTVRV 120

Query: 125 VENAEARLEAAL 136
 +EN ARLR+A+
 Sbjct: 121 LENGARLRSAI 132

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3357> which encodes the amino acid sequence <SEQ ID 3358>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2569 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/132 (88%), Positives = 128/132 (96%)

Query: 5 MIKIYTISSCTSCCKAKTWLNAHQLPYKEQNLGKESLTRDEILEILTKTESGIESIVSSK 64
 MIKIYTISSCTSCCKAKTWLNAH+L YKRONLGKE I/T++EIL IL+KTE+G+ESIVSSK
 Sbjct: 1 MIKIYTISSCTSCCKAKTWLNAHKLAYKEQNLGKEPLTKEILAILSKTENGVESIVSSK 60

Query: 65 NRYAKALNCNIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLERSIRN 124
 NRYAKAL+C+IEELSV+EVIDLIQ+NPRIKSPILIDDKRLQVGYKEDDIRAFLERSIRN
 Sbjct: 61 NRYAKALDCIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLERSIRN 120

Query: 125 VENAEARLEAAL 136
 +EN EARLEAAL
 Sbjct: 121 IENTEARLEAAL 132

Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 1088

A DNA sequence (GBSx1163) was identified in *S.agalactiae* <SEQ ID 3359> which encodes the amino acid sequence <SEQ ID 3360>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3097 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04987 GB:AP001511 unknown [Bacillus halodurans]
 Identities = 49/82 (59%), Positives = 64/82 (77%), Gaps = 1/82 (1%)

Query: 1 MGFDTETVRFRIIDGN-KVEISETLTAVYRSLERKGYNPINQIVGYLSGDPAFYPTIND 59
 M D T++F +++ V++ E L +VY +LEEKGYNPINQIVGY+LSGDPAFY+PR+ D
 Sbjct: 1 MSSNDNTNKFNVNEEPVSDVQEVLMGVYRALERKGYNPINQIVGYLSGDPAFYPRHKD 60

-1217-

Query: 60 ARNQIRKYRDEIVEELVRYYL 81
 AR IRK ERDE++EELV+ YL
 Sbjct: 61 ARTLIRKLERDEIVEELVKSYL 82

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3361> which encodes the amino acid sequence <SEQ ID 3362>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3097 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 80/88 (90%), Positives = 85/88 (95%)

- Query: 1 MGFTDETVRFRLLDGNKVGISETLAVYRSLKNGNFINQIVGVLSGDPAYVPRYND 60
 MGFTDETVRFR+LDD +K +ISETLAVY SL+KNGNFINQIVGVLSGDPAYVPRYND
 20 Sbjct: 1 MGFTDETVRFKLLDGDGRKISETLAVVHSLDKNGNFINQIVGVLSGDPAYVPRYND 60
 Query: 61 RNQIRKYRDEIVEELVRYFLQNGSID 88
 RNQIRKYRDEIVEELVRYFLQNGID+
 Sbjct: 61 RNQIRKYRDEIVEELVRYFLQNGIDV 88

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1089

- A DNA sequence (GBSx1164) was identified in *S.agalactiae* <SEQ ID 3363> which encodes the amino acid sequence <SEQ ID 3364>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1575 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10257> which encodes amino acid sequence <SEQ ID 10258> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CA314698 GB:299118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 82/138 (59%), Positives = 109/138 (78%), Gaps = 1/138 (6%)
 45 Query: 1 MRIMGLDVGSRTVGVALSDEPLGFTAGGLEIIKIDBESGNFPQFDRLAEVLKVKYKDKVFG 60
 MRI+GLD+G+KI+GVA+SD +G+TAQG+E IKI+E G++G RL+EL+K+Y +DK V+G
 Sbjct: 1 MRILGLDLGKTKLGVALSDEGNMTAQGIETIKINEAKGNYGLSRLSELIKDYTTIDKIVLS 60
 50 Query: 61 LPIKWNNTSGPRVEASQAYGDKITELNLPVVEYODERLTTVQARMLNVRQADISGRKKK 120
 PIKNWN T GPR EASQ + + +N+PV DERLTT+ AR+MI+ AD+SR KRKK
 Sbjct: 61 PPKNMGTVGPRGEASQTFYAKVLETTTNPVVLWDERLTTWAAREKMLI -AADVSRQGRKK 119
 Query: 121 VIDKLAACILQLNYLDRM 138
 VIDK+AA +IIQ YLD +
 55 Sbjct: 120 VIDKMAAVKILQGYLDSL 137

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3365> which encodes the amino acid sequence <SEQ ID 3366>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1575 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 114/139 (82%), Positives = 126/139 (90%)

Query: 1  MRIMGLDVGSKTIVGVAISDPLGFTAQGLEIIKIDERSGNFGFDRIAKLVKEYKVDKPVVG 60
      MRIMGLDVGSKTIVGVAISDPLGFTAQGLEIIKIDEE  FGF RL ELVK+Y+V++PV+G
Sbjct: 1  MRIMGLDVGSKTIVGVAISDPLGFTAQGLEIIKIDEEKARFGFTLRELVKQYQVEQFVIG 60

Query: 61  LPKQNMNTTSGPRVEASQAYGDKITELFNLPVEYQDERLITTVQERMLVQADISRGRKKK 120
      LPKQNMNT+GPRV+AS  YG+ I  LF  LPV  YQDERLITTV+A+RML+EQADISRGRKKK
Sbjct: 61  LPKQNMNTNGPRVDASITGYNGHIEHLFGLPVHYQDERLITVQERMLVQADISRGRKKK 120

Query: 121 VIDKLAQAQLILQNYLDRMF 139
      VIDKLAQAQLILQNYL+R  F
Sbjct: 121 VIDKLAQAQLILQNYLNRNF 139
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1090

A DNA sequence (GBSx1165) was identified in *S.agalactiae* <SEQ ID 3367> which encodes the amino acid sequence <SEQ ID 3368>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2631 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14697 GB:Z99118 yrzB [Bacillus subtilis]
Identities = 50/94 (53%), Positives = 65/94 (68%), Gaps = 5/94 (5%)

Query: 12  EHQHEVITLVDENQNETLFEILLITIDGRPEFGKNYVLLVPAGAEDDQGEITETQAYSFTE 71
      EH + IT+VD+ GNE L E+L T +  EFGK+YVL  P +++DE E+EI A SFT
Sbjct: 2  EHQENKNTITVDQGNELCEVLFTFEN-EPGKSYVLVYPIESKDDE--EVEILASSFTP 56

Query: 72  NADGTREGDLQPIPEDSDAEWMDIEEVPKSPLEDE 105
      N DG  G+L PI  ++D EWDMDIEE N+FL +E
Sbjct: 59  NEDEGENELFTI--ETDEEWDMDIEETLNTFLADE 90
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3369> which encodes the amino acid sequence <SEQ ID 3370>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3170 (Affirmative) < succ>
```


bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 90/98 (91%), Positives = 94/98 (95%)

Query: 7 HHDHNEHQHEVITLVDENGNETLFEILLTIDGREEFGKNYVLLVPAGAREDEQGEIEIQ 66
 H+H ++HQHEVITLVDE GNETLFEILLTIDGREEFGKNYVLLVPAG+EEDE GEIEIQ 66
 10 Sbjct: 3 HHDHNEHQHEVITLVDEQGNETLFEILLTIDGREEFGKNYVLLVPAGSERDRSGEIRIQ 62

Query: 67 YSFTENADGTEGDLQPIPEDSDAKWDMIEEVNSFLDE 104
 YSFTEN DTEGDLQPIPEDSDAKWDMIEEVNSFLDE
 Sbjct: 63 YSFTENSDGTEGDLQPIPEDSDAKWDMIEEVNSFLDE 100

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1091

A DNA sequence (GBSx1166) was identified in *S.agalactiae* <SEQ ID 3371> which encodes the amino acid sequence <SEQ ID 3372>. Analysis of this protein sequence reveals the following:

20 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2059 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1092

A DNA sequence (GBSx1167) was identified in *S.agalactiae* <SEQ ID 3373> which encodes the amino acid sequence <SEQ ID 3374>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

35 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -9.18 Transmembrane 314 - 330 (308 - 334)
 INTEGRAL Likelihood = -6.21 Transmembrane 279 - 295 (274 - 300)
 INTEGRAL Likelihood = -6.10 Transmembrane 136 - 152 (135 - 157)
 INTEGRAL Likelihood = -5.31 Transmembrane 232 - 248 (226 - 253)
 INTEGRAL Likelihood = -4.73 Transmembrane 163 - 179 (162 - 180)
 INTEGRAL Likelihood = -3.13 Transmembrane 95 - 111 (94 - 119)
 INTEGRAL Likelihood = -3.03 Transmembrane 386 - 402 (386 - 405)
 45 INTEGRAL Likelihood = -2.18 Transmembrane 204 - 220 (204 - 221)
 INTEGRAL Likelihood = -2.13 Transmembrane 40 - 56 (40 - 57)
 INTEGRAL Likelihood = -1.70 Transmembrane 186 - 202 (182 - 202)

----- Final Results -----

50 bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1220-

A related GBS nucleic acid sequence <SEQ ID 10255> which encodes amino acid sequence <SEQ ID 10256> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3375> which encodes the amino acid sequence <SEQ ID 3376>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.38    Transmembrane    315 - 331 ( 311 - 333)
      INTEGRAL    Likelihood = -6.48    Transmembrane    40 - 56 ( 37 - 61)
10   INTEGRAL    Likelihood = -6.10    Transmembrane    278 - 294 ( 274 - 298)
      INTEGRAL    Likelihood = -5.57    Transmembrane    392 - 408 ( 387 - 410)
      INTEGRAL    Likelihood = -3.98    Transmembrane    186 - 202 ( 184 - 208)
      INTEGRAL    Likelihood = -3.93    Transmembrane    339 - 355 ( 338 - 356)
      INTEGRAL    Likelihood = -2.97    Transmembrane    235 - 251 ( 228 - 253)
      INTEGRAL    Likelihood = -2.44    Transmembrane    166 - 182 ( 165 - 182)
15   INTEGRAL    Likelihood = -2.23    Transmembrane    106 - 122 ( 105 - 125)
      INTEGRAL    Likelihood = -1.81    Transmembrane    83 - 99 ( 83 - 101)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9179> which encodes the amino acid sequence <SEQ ID 9180>. Analysis of this protein sequence reveals the following:

```

25   Possible cleavage site: 13
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.38    Transmembrane    243 - 259 ( 239 - 261)
      INTEGRAL    Likelihood = -6.10    Transmembrane    206 - 222 ( 202 - 226)
30   INTEGRAL    Likelihood = -5.57    Transmembrane    320 - 336 ( 315 - 338)
      INTEGRAL    Likelihood = -3.98    Transmembrane    114 - 130 ( 112 - 136)
      INTEGRAL    Likelihood = -3.93    Transmembrane    267 - 283 ( 266 - 284)
      INTEGRAL    Likelihood = -2.97    Transmembrane    163 - 179 ( 156 - 181)
      INTEGRAL    Likelihood = -2.44    Transmembrane    94 - 110 ( 94 - 110)
35   INTEGRAL    Likelihood = -2.23    Transmembrane    34 - 50 ( 34 - 53)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.395(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 200/480 (41%), Positives = 310/480 (63%), Gaps = 1/480 (0%)

```

45   Query: 40  ILLYSVLSTLLAIANPLITYFANGLQTNLYTLGMLMTEQIQPYGDFVATGGFLYVTTIAL 99
      +L+S++ + L IA P LT ANGLQ+ONLY G+M+*KQQ+PYS F TOG Y+V IAL
      Sbjct: 40  LLFPTIIISLTIAPVPLTDRANGLQSONLYIGMMITKQGLPYSAAPTGTGGLFFVFTIAL 99

    Query: 100 SYLLGSSIWLLIVQFIATYVSGIYFYKLVIYVVAQSEIVSIQMTLFIYNNIVLFGFGMYF 159
      SY LGS+WI+ VQ +Y+SG+Y YKL+ Y+ + V++ ++ +Y++++ LGFGG+YF
50   Sbjct: 100 SYTLGSTLWLVNVPVQPCFPLSGLYLXLKLNMYTGPKQVALTPSISYTLGSLGFGGLYF 159

    Query: 160 IQWALFPMILISLWFLIKFCVDNIVDEAFIFYGILAAFLSIDFCQLILPWLCFVILATIN 219
      Q A+PP+LIS WFL K+ + DEAFI +G + A ++ IDP TLIFW + V + + N
55   Sbjct: 160 TQLAMPFLILISAWFLTKYFACLVKDEAFILFGFVGALAMLIDPSTLIFWSPACTVFGYN 219

    Query: 220 IQKQKSLRGFYQFLCVFVGMILLIATYVGYFMFNLIQISSYIDKAIIFYPPTYFARTNHSFL 279
      I QK RGFYQ L +FGMIL+ YT GYP+ NLQ++ Y+ + YPPT+ F N S FL
      Sbjct: 220 ISQKHLARGFYQLLASIFGMILVFYTAGFYILNLQVLINFLYLSQTMYYPTFFKSGSLGSL 279

60   Query: 280 LSLAIQIVVLLGGSCFLGLNDPFIQNRKKASYQIGLNFIACTPIYVAINAIFSRDNLFYHF 339
      LAIQ+ LG G L G+ + I+ K S ++ + + +AIFG+D+ YH
      Sbjct: 280 FGLAIQLFALGLGLLGMENVIRKPKNSDRVVKVLFVNVILESILVAIFGQYRPHL 339

```

-1221-

5

Query: 390 LPALPFGLLT/TSNKITLYLVQKIDRRSHQYR-FGSKLSVLLPFLYLPPLLWLSLSI 398
LP LPFG+L+T+ + Y + + SERR+ +G + + + + + +YLP+L+V +
Sbjct: 394 LPLLPFGLLTAIAPVQYVQIGLGSGSERRRRRKGKVGKRWIMLYKRRFYLPLLVGTILL 399

10

Query: 399 GLIVNTYQNVLTKSKRDISHYLTIKLRDGKIVLWQKVASIYQZRLKSAQVPLFI 458
Y + L+ + + + + YL K+ + + + + IVWD + IY + + KS SQF P +
Sbjct: 400 ICTCFQNISSFLAQGRDHRTAAYLEQIKLNTQSYIVWDVTKYVLSKSAQVQSSPDI 459

Query: 459 NTAQSSIKRLKDELLGQAKYKPLNNRKLNNELKSLDKKHYYVLSNITHVLYRPFK 468
NT + + + + KIL+DELL+L A X + + N+ + L + + + + +Y+ F+ +Y+ K
Sbjct: 460 NTKESIRKILKDELLGAKYIVNRYNRYKPLTKQVLSNITHVYKDEIITTSKYVOKK 519

A related GBS gene <SEQ ID 8727> and protein <SEQ ID 8728> were also identified. Analysis of this protein sequence reveals the following:

```

15 protein sequence reveals the following:
    Lipop: Possible site: -1  Crend: 0
    SRCLFG: 0
    MoG: Length of UR: 34
        Peak Value of UR: 2.23
20    Net Charge of CR: 0
    MoG: Discrim Score: 7.72
    GVH: Signal Score (-7.5): -2.21
        Possible site: 60
    >>> Seems to have a cleavable N-term signal seq.
25    Amino Acid Composition: calculated from 61
    ALCOM program count: 5 value: -9.18 threshold: 0.0
        INTEGRAL Likelihood = -9.18 Transmembrane 174 - 190 ( 168 - 194)
        INTEGRAL Likelihood = -6.21 Transmembrane 139 - 155 ( 134 - 160)
        INTEGRAL Likelihood = -5.31 Transmembrane 92 - 108 ( 86 - 113)
30    INTEGRAL Likelihood = -3.03 Transmembrane 246 - 262 ( 246 - 265)
        INTEGRAL Likelihood = -2.18 Transmembrane 64 - 80 ( 64 - 81)
        PERIPHERAL Likelihood = 3.29 194
    modified ALCOM score: 2.34
    icml HYPID: 7 CFP: 0.467
35
    *** Reasoning Step: 3

    ----- Final Results -----
        bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>
40        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF02392(331 - 978 of 1764)
45 EGAD[43696][MJ1079(2 - 379 of 397) conserved hypothetical protein [Methanococcus jannaschii];
GMNI[MJ1079 conserved hypothetical protein GP1591727[gblAAB99076.1]]U67550 conserved
hypothetical protein [Methanococcus jannaschii] PIR[P64434|P64434 hypothetical protein
MJ1079 - Methanococcus jannaschii
%Watch = 3.1
50 %Identity = 25.6 %Similarity = 50.7
Matches = 57 Mismatches = 100 Conservative Sub.s = 56

      174       204       234       264       294       324       354
55 *LLLANI*LSVHPTSFPTXXXN*LXXSSIWLLVQFYATYVSISYFKLYVTVQAKSVSIGMLIFYIMNVILG-----
                                           :||:||:||||
                                           MNLLYLILGLICSTTGL
                                           10

      426       447       477       507       537       567       597
60 FGGNYPIQWALPPMLISLWFL----IKFCVNVIDEAFIFGYIGLAAPSLIPDPTLIPLWCSEVLLITATNIKOKQSRLRPF
   ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
FPGIHPNNIVALSPLLIPYGELGNLYIFFLGLVIHYHFNF-IPSAFLGVPEDETAVALPMKHTLNNGNYETAVLAGP

           30         40         50         60         70         80         90

65    627        657        687        717        747        774

```

-1222-

```

YQFLCVVFPGLILAYTVGYFMPNLQIISSYIDKALFYPPPTYPARTNHSFLSLAI-QIVVLLGSGC-----
| | | | | : : : : : : : : : : | | | | : : : : : : : : : :
GSYLGVVSPSLILISLFLMSILHFDVRAFYCSI--KIPFPFLIAFILYQIPTAKSVWSEVLVFLPSGPIGLIAYVYCSEAFNI
110      120      130      140      150      160      170

5      798      828      846      876
-----LGLNDLPIONRKKASYQ-----IGLNPTIACIFI
| | | : | | :
TLTAITPFGMPSLFLNNLKYTIKISKQMAFDPDFELKPLKSGPFA----TIAITILNLSKYILLPIRKVNFKPLSLPFI
190      200      210      220      320      330

10      906      948      978      1008      1038      1068      1098
IYAIMAIFSRDFN---LYH--FLPALPFGLL/TSNKITILYQKVIDRSHRRQYFSGKSLIVDLFVKKIYYLPLLLVSL
| : : : : | : | : | : | : | :
15      IPCSLVVIIGSYNTYLYHILVYLTAIYIGLAVKSNLNSMNVLIPTIPLYFLRG
350      360      370      380      390

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1093

A DNA sequence (GBSx1168) was identified in *S. agalactiae* <SEQ ID 3377> which encodes the amino acid sequence <SEQ ID 3378>. This protein is predicted to be anaerobic ribonucleotide reductase (nrdD). Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3722 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10253> which encodes amino acid sequence <SEQ ID 10254> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

35 >GP:AAD00215 GB:U73336 anaerobic ribonucleotide reductase
[Lactococcus lactis subsp. cremoris]
Identities = 539/725 (74%), Positives = 616/725 (84%), Gaps = 7/725 (0%)

Query: 10 NTESDIKVIKRDGRVFPDKYKIYATALLKASNKVTKMSPLVEAKLMIADHVIAEINYRF 69
+T +I VIKRDGR V F+ KI+ AL KA+ KV V L + D V++EI+ +RF
40 Sbjct: 10 VTLEENIVIKRDGRSVKFNSEKIFDALTAKKAKVELTKGV---LEELTRVVEI+SRF 66

Query: 70 KDKIKIYIQLNIVHKLKLANEYATQAQYINRYNRTQDFERSQATDINFSIGKLINQDQTV 129
+N+KIYELQ+IVE +LLE+ E A+A+EYI+YR RD R+ATDINF+I KLIN+DQTV
45 Sbjct: 67 SENVKIYELQISVBQELLESGETALAREYISYRANREDLARTKATDINTPTKELINRDQTV 126

Query: 130 VNENANKSDVFNTRDRLTAGIVGKSGIGLQMLPSHVANAHQKGDHYHDLDSYPTMTN 189
VNENANKDS+VFNTQRDLTAG V K+IGLK+LP HVANAHQKGDHYHDLDSYPT M N
Sbjct: 127 VNENANKDSNVFNTRDRLTAGAVSKAIGLKLLPPHVANAHQKGDHYHDLDSYPTMTN 186

50 Query: 190 CCLIDFKMGLANGPKFIGNAEVSPKSIQTATQAISQIIANVASSYGGCTADRIDEFLAP 249
CCLIDFK M NGPK+GNA+V+SPKSIQTATQA SIIANVASSYGGC+ DR DE LAP
Sbjct: 187 CCLIDFKMFPNGPKFIGNAQVDSPKSIQTATQAQSIANVASSYGGCSFDRADEVFLAP 246

55 Query: 250 YAQNLNQKHLKDAKSWVIED-KQEDYARAKTQKDIYDAMQSLVEYINTLFTSNQGTPTFS 308
YA+LNYQKHLKDA+ +W+ D K+E YAR KT KDIYDAMQSLVEYINTLFTSNQGTPTF +
Sbjct: 247 YAKLNYQKHLKDAQINDGDEKREYARAKTAKDIYDAMQSLVEYINTLFTSNQGTPTFV 306

Query: 309 LGFGLTNWPFERIKALKIRIQGLGSEHRTAIFPKLIFLTKKGLNLEEDSPNYDIKQL 368

```

```

+GFGIG +W+ REIQKALK+RI GLGSEHRTAIPFKLIPTLK+GLNLE +PNYDIK+L
Sbjct: 307 VGFGLGDDWYAREIQKALKIVRIIGLGSEHRTAIPFKLIPTLKRIALLEVQTPNPDIKEL 366

Query: 369 ALBCATKRMYPDVLSDYDKI+DLTGSFKAPMGCRSFLQWRDANGQVTSGRNNLGVVTVN 428
ALBC+TKRMYPD+LSYDKI++LTGSFKA MGRSFLQW+DANG DVT+GR NLGVVTVN
Sbjct: 367 ALBCSTKRMYPDILSDYDKIVELTGSFKASMGCRSFLQWRDANGNDVTSGRNNLGVVTVN 426

Query: 429 LPRVAMESNGDMDFWEI FNERMSIARDALVYRVERVKEAIPANAPILVYQYAGPRLGK 488
LPR+A+E+ G+ +KFWEI FNER+ IA DAL +RVER KEA P NAPIL+ GA G RL
Sbjct: 427 LPRIALAAGNKEKFWEI FNERVETIAHUALPRVERAKEAQKNAPILFNGGALG+RLDS 485

Query: 489 YDNVDRLFNHRRATVSLGYIGLYEVA+VYFGDWNEDNEQAKAFTVDIVRKMQLCADWSD 548
+VD L+N+ RATVSLGYIGLYEVA+ FYG WE N +AKAFT++IV++M + C DWS
Sbjct: 486 EGSVDDLNNERATVSLGYIGLYEVAFTFYGPTWESNPEAKAFTIEIVKRMRHDCEDWSK 545

Query: 549 EYDYHFSVYSTPSSSLTORPCRLDTEKFGIVDITDKEYTINSFHYDVKNPTPFKEKLF 608
YH+SVYSTPSSSLTORPCR+D EKFG V DITDE+YTTNSFHYDVKNPTPFKEKLF+
Sbjct: 546 ASGYHYSVYSTPSSSLTORPCRMDKEKFGSVADITDKDYTNSFHYDVKNPTPFKEKLF 605

Query: 609 EKLYPETGASGGFIHYCHYFVLQGNPKALSAVNDYADRVGYLGTNPIDKCYCQCFEGD 668
EK YP A+GGFIHYCEYFVLQGNPKALSAVND+AYDR+GYLTN PID CY C FEGD
Sbjct: 606 EKDYF-VYANGGFIHYCEYFVLQGNPKALSAVNDYADRIYGLTNPIDHCYACGFE 664

Query: 669 FTPTDRGFCTPCNCGNDPKTDVVVKRTCSYLGNPQARPMVNGRHKEL SARVKHMGNS-SI 727
FTPT+RGF CP CGN DPKT DVVKRTCSYLGNPQARPMV+GRHKEL S+RVKHMGNS
Sbjct: 665 FTPTERGFCQCGNDPKTDVVVKRTCSYLGNPQARPMVNGRHKELSSRVKHMGNSVGA 724

Query: 728 KNQGN 732
N GN
Sbjct: 725 LNDGN 729

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3379> which encodes the amino acid sequence <SEQ ID 3380>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2975 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 641/731 (87%), Positives = 680/731 (92%)

Query: 1 MVLRLRERFMTESDIKVIKEDGRIVSFDKYKIYTALLKASNKVIRMSPLVEAKLEMIADH 60
M+ LE ++ + DIKVIKEDGRIV+FD KIY+ALLKAS KV +MSPLVEAKLE I+D
Sbjct: 1 MVLSEDKVTVPDVIKVIKEDGRIVNFDSTKIYSALLKASKVITMSPLVEAKLEAISDR 60

Query: 61 VIAETYNRPKDNKIYVEIQNIVEHKLAEANEYALAEQYINRYTQRDPERSQATDINFSIG 120
+IAEI RP NIKIYRIQNIIVEHKIL ANKYALA+EYINRYTQRDP RSQATDINFSI
Sbjct: 61 ITAEIIEFPPTNIKIYRIQNIIVEHKLAEANEYALAEQYINRYTQRDPERSQATDINFSID 120

Query: 121 KLINKQQTVVNENANKISDVFPNTQRDLTAGIVGKSIGLJMLPSHVANAHQGDIIHYHLD 180
KLINKQQTVVNENANKISDVFPNTQRDLTAGIVGKSIGLJMLPSHVANAHQGDIIHYHLD
Sbjct: 121 KLINKQQTVVNENANKISDVFPNTQRDLTAGIVGKSIGLJMLPSHVANAHQGDIIHYHLD 180

Query: 181 YSPYTPMTNOCCLIDPKGLANGFKIGNARVESPKSIQTATAQISQIIANVASSQYGGCTA 240
YSPYTPMTNOCCLIDPKGLANGFKIGNARVESPKSIQTATAQISQIIANVASSQYGGCTA
Sbjct: 181 YSPYTPMTNOCCLIDPKGLANGFKIGNARVESPKSIQTATAQISQIIANVASSQYGGCTA 240

Query: 241 DRIDEFLAPYAQLNYQKHLDAKSWIEDKQEDYARAKTQKDIYAMQSLYEINTLPTS 300
DRIDEFLAPYA+LN++KH+ DAK+H++E K+E YA KTQKDIYAMQSLYEINTLPTS
Sbjct: 241 DRIDEFLAPYAKLNKKHMAKAKWIVETKRESYAFKTKQDIYAMQSLYEINTLPTS 300

Query: 301 NQQTPTSLGFLGNTWPFEREIQKALKIRIQGLSEHRTAIPFKLIPTLKGLNLEEDS 360

```

-1224-

NGQTPFTSLGPGIGT-WFEREIQKAIL IRI GLSEHRTAIFPKLIPT+K+GLNLE DS
 Sbjct: 301 NGQTPFTSLGPGIGTISWFEREIQKAIL/TIRINGLGSEHRTAIFPKLIPTVKRGLNLEPDS 360
 Query: 361 PNYDIKQLALECATKRMYPDVLISYDKIIDL/GSPFK+PMGCRSFLQGWDRANGQDVTSGRM 420
 PNYDIK LALECATKRMYPD+LSYDKIIDL/GSPFK+PMGCRSFLQGW+D NQDQDVTSGRM
 Sbjct: 361 PNYDIKTLALECATKRMYPDMLISYDKIIDL/GSPFK+PMGCRSFLQGWDRANGQDVTSGRM 420
 Query: 421 NLGVVTVNLPRVAMESNGMDKFWRIINERPMSTARDLVRVVRVKEAI PANAPILYQYG 480
 NLGVVT+NLPR+AMESNGMDKFWRIINER I++DAL+YRVVRV EA PANAPILYQYG
 Sbjct: 421 NLGVVTLNLPRAMESNGMDKFWRIINERMLISKDALIRVVRVTEAKPANAPILYQYG 480
 Query: 481 AFGERLKGYNVDRLFNIRRAIVSLGYIGLYEVASVFGDMEDNHQAKFTVDIVRMK 540
 APG+RL K NV+ LF +RRATVSLGYIGLYEVASVFGG WR N AKAFI+ IV+ MK
 Sbjct: 481 APGRLEKTGNVNLFKMRRAIVSLGYIGLYEVASVFGGQWEGNDAKAFITLIVKMK 540
 Query: 541 QLCADNSDEYDVHFSVYSTPSESLTDRFCRLDTEKPGIVIDITDKEYYTNPSHYDVRK 600
 Q C DNSDEY YHFSVYSTPSESLTDRFCRLDTEKPGIVIDITDKEYYTNPSHYDVRK+P
 Sbjct: 541 QACADNSDEYGVHFSVYSTPSESLTDRFCRLDTEKPGIVIDITDKEYYTNPSHYDVRK 600
 Query: 601 TPFEKLDPEKIYPTGASGGPIHYCEYFVLQNPKALEAVMDYADRVGLTNTPTIDKC 660
 TPFEKLDPEK YPE GASGGPIHYCEYFVLQNPKALEAVMDYADRVGLTNTPTIDKC
 Sbjct: 601 TPFEKLDPEKIYPTGASGGPIHYCEYFVLQNPKALEAVMDYADRVGLTNTPTIDKC 660
 Query: 661 YQCQFEGDFTPTDRGFTCPNCGNSDPKTVDVVKRTQCYLGNPQARPMVNGRHKRISARVK 720
 Y QCQFEGDFTPT+RGFTCPNCGN+DPKTVDVVKRTQCYLGNPQARPMVNGRHKRISARVK
 Sbjct: 661 YNCQFEGDFTPTDRGFTCPNCGNSDPKTVDVVKRTQCYLGNPQARPMVNGRHKRISARVK 720
 Query: 721 HMGSGSIKNGQ 731
 HMGSG+IK G
 Sbjct: 721 HMGSGSIKNGQ 731

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1094

35 A DNA sequence (GBSx1169) was identified in *S. agalactiae* <SEQ ID 3381> which encodes the amino acid sequence <SEQ ID 3382>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5372 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3383> which encodes the amino acid sequence <SEQ ID 3384>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6084 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 28/47 (59%), Positives = 40/47 (84%), Gaps = 1/47 (2%)

Query: 1 MGKTYLADYKQVQVQKPEHSHSTGEMANQKSLKIDLRKQPLEKAKKK 47

-1225-

MGKYQLDYKG QV++PHEKHS + ++KSR++L+ +PLEK+KK+
 Sbjct: 1 MGKYQLDYKMQOQVERPHEKHSK-KTKDKSRVQELKARFLEKSKKO 46

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1095

A DNA sequence (GBSx1170) was identified in *S. agalactiae* <SEQ ID 3385> which encodes the amino acid sequence <SEQ ID 3386>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0436 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE95794 GS:AL359949 putative oxidoreductase [Streptomyces
 coelicolor A3(2)]
 Identities = 91/299 (30%), Positives = 147/299 (48%), Gaps = 7/299 (2%)

Query: 2 IQLGIVGLGKISQKAYLPFMRQVTGVVHHLFTROKQILREV--NMLFGSSTAYDSISLA 59
 +++G+GLG I+QK YLE + + G+ HL TR L V + + + LD+L
 Sbjct: 1 MKVCGTIGLADIAQKGLYPLAALPGI EHLQITRTFATLTVADKLRIPPPQRRHADLALL 60

Query: 60 EHPHLDGVTHVATSAHFQIAKLPVFMKPLTEDYTSTKALYDLAKDHKTFLMG 119
 ID F+H T+H+ I L+ G+P ++DKPL + ++ L L+ + T L G
 Sbjct: 61 AQLDAAFPVHAPTAHFEIVITLLEAGVPTTYDKPLAYLEADSERLVTLAEERGTSILAV 120

Query: 120 FMRERFAPRINEMKKVSDQHLRTFPRANVAFADFOYKLFDMFPHLDTALFLTNVVRG 179
 FMRR AP + + + I KN P D + + D FTH +DT FL V
 Sbjct: 121 FMRRHAPGYAQCAB-HPRELILMQNRTGLPEDPRTMILDDFIHVVDLTFPLVPGVDDV 179

Query: 180 YFVTKRDGNKILQVSVILETDSIEIIEASMLQSGSRRRIIEIESPEVTVSLDDLSMLSVI 239
 + + G + V + L D MN SGS EIE+ + + + L+ VI
 Sbjct: 180 TVRARTSGCLLHFVVLQAGDQFTALGVNRLSGSAGEILEVSGQDTKRVNLA--EVI 237

Query: 240 DGFDRRAI-FGGSWASTLEKRGFEPMDIAFIQAITGVNPIPSKSLLSHFICQIQINKA 297
 D + + G W +RG E + AF+ A+ +G +S +L +H +C++ +A
 Sbjct: 238 DHKQPTVRRRGDMVPVARGGIEQVLAFLDAVRSG-EVLSARDALATHELCSRVRA 295

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3387> which encodes the amino acid sequence <SEQ ID 3388>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF96942 GS:AE004430 oxidoreductase, Gfo/Idh/MocA family [Vibrio cholerae]
 Identities = 103/304 (33%), Positives = 158/304 (51%), Gaps = 11/304 (3%)

Query: 4 INIGIVGLGAIQKAYLPYMRQLSDITWHLSSTRNAVRQVQGLPGHALLYSIDVKELKKT 63
 + I +GLG I+QKAYLP + Q DI L TRN V + + + +D +++ +
 Sbjct: 1 MKIANIIGLADIAQKAYLPVLAQWFDIELVLCTRNPKVLGTATRYRSATCTDVRDLQY 60

-1226-

Query: 64 NLDGVFIHAATSAHAELASFLAQGIPVFMKPTADNYLTKNLYDLAKENQTFIMAGFN 123
 +D V IHAAT H+ LA+ FL+ GIP F+DKP+A + +NLV+LA+++ L GFN
 Sbjct: 61 GVDVMIHAATDVHSTLAAFLFHLGIPTFVVDKPLAASQCEENLYELAEKHQPLVVGFN 120

5 Query: 124 RRPTPRVKK-LSSLSTK-----RKVAVEKNDLARPQDMTFKLPDPFFIHLDTALFLTEGT 177
 RR P + LS L+ + R + EK+ FQD+ +FD FIHPLD+ +
 Sbjct: 121 RRHIFLYNQHLSELAQQECQALRSFLRWKHKHRLPGDIRTFVFDFFIHLDSVNLRSQCN 180

10 Query: 178 LLKGGPQVHLEAGLLSQVMVLTMTESMTTASMLQSGSRREVMVQRAEETHYLENDEL 237
 L H YH+ QLL+++ V T ASMN QG E + Y ++ +
 Sbjct: 181 LDDLHLTYHMSBGLLARLDVQWQTGDTLLHASMNRQPGITTEHVASYDENVAYFDSPQTQ 240

15 Query: 238 LSIYKTEKRVLFQFASMDTTLHKRGFETMIDAFLEAISTGVNFPV-PSSLLSHW-----I 292
 +++ ++ + W L +GP+ M+ +L+ + G P E +L SH I
 Sbjct: 241 GKWRDQKQBSRVALKDWTMFLASKGFDAMVQWMLQVAAAGKLPHTHIERNLASHQLAEAI 300

Query: 293 CQOI 296
 CQOI
 Sbjct: 301 CQOI 304

20

An alignment of the GAS and GBS proteins is shown below.

Identities = 168/308 (54%), Positives = 223/308 (71%)

25 Query: 1 NLQGLGIVGLGGISQKAYLFYMRQVTGVVHFLFTRKQILEEVNMLFGSSTAYDSLDSLA 60
 NL +GIVGLG ISQKAYLFYMRQ++ + WHL TR + ++V LFG + Y + L++
 Sbjct: 3 MLNIGIVGLGAI SQKAYLFYMRQLSDITVHLSTRAAVRQVQQLFGHAILYSVVKLSK 62

30 Query: 61 HPLDGVFIHVATSAHFDIAKLFKKGIPVFMKPLTEDTSTKALYDLAKDHKTFLMAGF 120
 LDGVFIH ATSAH ++A LFL +GIPVFMKDP+ ++Y TK LYDLAK+++TFLMAGF
 Sbjct: 63 TNLDGVFIHAATSAHAELASFLAQGIPVFMKPTADNYLTKNLYDLAKENQTFIMAGF 122

35 Query: 121 NRRFPAPRIMEMKFKVEDGKHIRTFFQAVNAPADFYKLPFMPFIHLDTALFLTNVVRKG 180
 NRRF PR+ ++ + K + KH +N P D +KLF DTIHLDTALFLT + +G+
 Sbjct: 123 NRRPTPRVKKLSLSTKRVAVEKNDLARPQDMTFKLPDPFFIHLDTALFLTEGTLKGH 182

40 Query: 181 FVTIKDGNKILQVSVLETDSEIIRASMNQSGSRREIIRIESPEVTVSLDLSNLVID 240
 F + + QV VIL T+S ASMNQSGSRRE++E++ E TY L++L L6+
 Sbjct: 183 FQYHLEAGLLSQVMVLTMTESMTTASMLQSGSRREVMVQRAEETHYLENDELSEIYK 242

45 Query: 241 GFDRRAIGFGSNASTLEKRGFEMIDAFITGQVNPISPKSSLSHPICDQINKANAP 300
 G ++R +GF SW +TL KRGFE MIDAF++AI+TGVPN+SP+SSLSH+IC QI +
 Sbjct: 243 GTEKRVLFQFASMDTTLHKRGFETMIDAFLEAISTGVNFPV-PSSLLSHWICQIADSQLS 302

Query: 301 FGMLNLKI 308
 +G L +++
 Sbjct: 303 YGELTVEL 310

SEQ ID 3386 (GBS309) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 10; MW 63kDa).

50 GBS309-GST was purified as shown in Figure 212, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1096

55 A DNA sequence (GBSx1171) was identified in *S. agalactiae* <SEQ ID 3389> which encodes the amino acid sequence <SEQ ID 3390>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1227-

bacterial cytoplasm --- Certainty=0.2983 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAE04222 GB:AP001508 unknown conserved protein in others
 [Bacillus halodurans]
 Identities = 52/129 (40%), Positives = 70/129 (53%), Gaps = 5/129 (3%)

10 Query: 39 FEDWLDENLNMELGVGVDPNFVPIQVVSFNDNNAIGFLNLRRLNDTLLEKGGHIGYS 98
 FE L ++ GV +P N V + IG +N+R LND L +GGHIGY
 Sbjct: 43 FEHLATLKDYQHGVNLPANRVAMTTMLVHEBQKRLIGAINIRHTLNDWLHHRGGHIGY 102

15 Query: 99 IRPRGRGKYAKBQLKLSIEQAHLKNINSILVTCVNDNASKSVILANGVLEDCILHQ-- 156
 IRP +RGKGYA LKLS+H+A ++L+TC 4N S I HGGVLA+ +
 Sbjct: 103 IRPSERGGKYATLMLKLGLERAAALGLEKVLITCDKENLPSARTICRGGGVLDGEVVDIER 162

Query: 157 ---TERIWI 162
 +RIWI
 20 Sbjct: 163 GIAIQRWI 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3391> which encodes the amino acid sequence <SEQ ID 3392>. Analysis of this protein sequence reveals the following:

Possible site: 59
 25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2195 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 90/164 (54%), Positives = 115/164 (69%), Gaps = 4/164 (2%)

35 Query: 1 MKLRRFVLEDKEEILAMNKEFQKESSSDVG--GFYEPTMHFELMDLNLNMLGVGVDPN 59
 M++RRP L+DK+ +L+M EF ++ S+ DG F ++E WL+ +L E+G+
 Sbjct: 1 MKIRRPPLKDKDAVLSEMINFELEQKSAITGLWHFPNVNDFNYETWLEDLSLRQEMGLS--SQ 58

40 Query: 59 FVPIYIQVVSFNDNNAIGFLNLRRLNDTLLEKGGHIGYSIRPRGRGKYAKBQLKLSIE 118
 VP IQ+V+FD + AIGFLNLRRLN+ LLEKGGHIGYS+RP GRGKYAKB LK +
 Sbjct: 59 GVPAIQVAVFDRSQAIQFLNLRRLNRLNRLLEKGGHIGYSVRFSGRGKYAKBMLKQAVS 118

Query: 119 QARLANINEILVTCVNDNASKSVILANGVLEDCILHQTERIWI 162
 A KNI ILVTC N HSA++VI+AN G+LEL TERIWI
 45 Sbjct: 119 YALSNITITLVTCDETNVASRAVIVANVILEDSRGCTERYWI 162

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1097

50 A DNA sequence (GBSx1172) was identified in *S.agalactiae* <SEQ ID 3393> which encodes the amino acid sequence <SEQ ID 3394>. This protein is predicted to be anaerobic ribonucleotide reductase activator protein (nrdG). Analysis of this protein sequence reveals the following:

Possible site: 59
 55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4239 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1228-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 5 >GP: AAD00216 GB: U73336 anaerobic ribonucleotide reductase activator
protein [Lactococcus lactis subsp. cremoris]
Identities = 152/198 (76%), Positives = 176/198 (88%)
- Query: 8 NTPKPGEMKSEELSHGHIIIDYKAFNFDGEGVRSNLYVAGCMFHCKGCYNATWSF 67
N PKPGEM++RLS +I DYK FNFVDGEGVR GLYV+GCMFHC+GCYN ATWSR G
- 10 Sbjct: 2 NNPKPGEMRAELSQNYIADYKPFNFDGEGVRSNLYVAGCMFHCKGCYNATWSF 61
- Query: 68 FYTKELEDQIMTDLEQPYVQGITLGGEPFLNTGILLPLLQIRRELPEKDIWSWTGYT 127
FYTKELED-IM DL +PYVQGITLGGEPFLNT L+PLL+RIRRELPAKDINSWTGYT
- 15 Sbjct: 62 FYTKELEDKIMADLAEPYVQGITLGGEPFLNTYIPLILKIRRELPAKDINSWTGYT 121
- Query: 128 EEMMLETPDKLEMLSLIDILVDGRFDQSKRNLMQFRGSSNQRIIDVQKSLKEGEVVIWE 187
EEMMLET DKLEML L+D+LVDGRF+ SK+NLMLQFRGSSNQRIIDV KS +G+VVIWE
- 122 EEMMLETDDKLEMLDLDLVDGRFELSKRNLMQFRGSSNQRIIDVPSRSKQVVIWE 181
- 20 Query: 188 KLMDGDNSTYQVKRDLL 205
LMDG+N+EQ++L+
- Sbjct: 182 KLMDGNNFQIHKEKLI 199

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3395> which encodes the amino acid
25 sequence <SEQ ID 3396>. Analysis of this protein sequence reveals the following:

- Possible site: 59
>>> Seems to have no N-terminal signal sequence
- 30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4111 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 35 Identities = 167/202 (82%), Positives = 186/202 (91%)
- Query: 4 EASHNTPKPGEMKSEELSHGHIIIDYKAFNFDGEGVRSNLYVAGCMFHCKGCYNATWSF 63
E WN PKP EM++EELS G IIDYKAFNFDGEGVRSNLYV+GCMFHCKGCYN ATWSF
- 40 Sbjct: 4 EKCMQNPKEPKEMQAEELSQRIIDYKAFNFDGEGVRSNLYVSGCLFHCCKGCYNATWSF 63
- Query: 64 RAGIFYTKELEDQIMTDLEQPYVQGITLGGEPFLNTGILLPLLQIRRELPEKDIWSWT 123
+AG+PYT+EL+QIMTDL QPYVQGITLGGEPFLNTGIL+PL++RIRRELPEKDIWSWT
- 64 Sbjct: 64 KAGMFPYTQLELEDQIMTDLAQPYVQGITLGGEPFLNTGILPLILKIRRELPEKDIWSWT 123
- 45 Query: 124 GYTWEEMMLETPDKLEMLSLIDILVDGRFDQSKRNLMQFRGSSNQRIIDVQKSLKEGEV 183
GYTWEEMMLET DKLEMLSLIDILVDGRFD +K+NLMLQFRGSSNQRIIDVQKSL EV
- 124 Sbjct: 124 GYTWEEMMLETPDKLEMLSLIDILVDGRFDITKRNLMQFRGSSNQRIIDVQKSLAKEV 183
- 50 Query: 184 VIWEKLMDGDNSTYQVKRDLL 205
+IW+ LNDGD++BQ+ R+DLL
- 184 Sbjct: 184 IIVDKLNDGDQTFQIGERDIL 205

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1098

A DNA sequence (GBSx1173) was identified in *S. agalactiae* <SEQ ID 3397> which encodes the amino acid sequence <SEQ ID 3398>. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

-1229-

INTEGRAL Likelihood = -3.03 Transmembrane 102 - 118 (101 - 119)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AD24446 GB:AF118389 unknown [Streptococcus suis]
 Identities = 97/240 (40%), Positives = 151/240 (62%), Gaps = 1/240 (0%)
- Query: 2 IKILIPAKEMKV-CQNIANPKLSAQTKIILDPSTLTVSLELDIYRINTSAARCEQRW 60
 +KI+IP AKE+ +N ++ LS ++K ++D S V + Y++N + A EA RW
 15 Sbjct: 1 NKIIIPNAKEVNTNLENASFYLLSDRSKFVLDASIQFDVKMAAFYKLNEAKALEADRW 60
- Query: 61 QDFKAKQLTLNPAIKLFLNGLMVRNIKRNLSTSEAQFMENSIVTTSALYGLIIPAMTLISP 120
 + Q PA +L+GLMYR + R + + E ++ + V + +ALYG+I ISP
 15 Sbjct: 61 YRINTQAKTYPAWQLYDGLMTRYMDRRGIDSKEENYLADHVRVATALYGLIHPPFETISP 120
- 20 Query: 121 HRLDFNTKIKINNSLKVFMRENVDYTFMQSDDIMVSLLSNEFETVPSPKERQKLILHINFI 180
 HRLDF +KI N SLK +NR YD + D+++SL S+EFE VFSP+ ++L+ + F+
 Sbjct: 121 HRLDPQSLKIGNQSLKQYMRPYDQEVGDDLELTLSSASEFQVFPQIQRLVKILFM 180
- 25 Query: 181 EDRDQQLKTHSTISKARGKCLTAMMENNCTLEHLKQLRFDGFCYDNELSDSKQLTFVK 240
 E++ GQLK HSTISK RG+ L+ + +NN Q L ++ + DGF Y S + QLT++
 Sbjct: 181 EEEAGQLKTHSTISKARGKCLTAMMENNCTLEHLKQLRFDGFCYDNELSDSKQLTFVK 240

A related GBS nucleic acid sequence <SEQ ID 10941> which encodes amino acid sequence <SEQ ID 10942> was also identified.

- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3399> which encodes the amino acid sequence <SEQ ID 3400>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3759(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 40 An alignment of the GAS and GBS proteins is shown below.

Identities = 114/242 (47%), Positives = 155/242 (63%)

- Query: 1 MIKILIPAKEMKVQNIANPKLSAQTKIILDPSTLTVSLELDIYRINTSAARCEQRW 60
 N+ LIPTAKEM + + L ++ I+ + +T DL YRI ++A E QRW
 45 Sbjct: 1 MLTFLIPTAKEMTIPKESHPIILLPQDSQAILKIMAMTTEDLAKSYRIKESAKKEQRW 60
- Query: 61 QDFKAKQLTLNPAIKLFLNGLMVRNIKRNLSTSEAQFMENSIVTTSALYGLIIPAMTLISP 120
 QD ++Q PA +L+ENGLMYR+IKR L+T E ++ V+ITS+ YGIIPA I+
 50 Sbjct: 61 QWASQSLAYPAYQLFNGLMYRHIKRDKILITQSAYLITQVYITSSFGIIFANHPHIAE 120
- Query: 121 HRLDFNTKIKINNSLKVFMRENVDYTFMQSDDIMVSLLSNEFETVPSPKERQKLILHINFI 180
 HR DF+T+IKI SLK +NR Y+ F + +SLLS+EF+ VFS +Q I L P+
 50 Sbjct: 121 HRLDFHTRIKIEQSLKSYMRPCYNQFAKEHFQVITSLSEFDDVFSEKDCQKQWISPKFM 180
- 55 Query: 181 EDRDQQLKTHSTISKARGKCLTAMMENNCTLEHLKQLRFDGFCYDNELSDSKQLTFVK 242
 +++GQ KTHSTISKARG LTA MENNCT++ LK L P GF Y +LS + ++KK+
 Sbjct: 181 AEKEGQPKTHSTISKARGAFLTACMENNCTVDSIKSLVFAGFYYPHLPSTHIEFYVIKKK 242

-1230-

SEQ ID 3398 (GBS428) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 6; MW 30.4kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 4; MW 55kDa).

GBS428-GST was purified as shown in Figure 220, lane 6-7.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1099

A DNA sequence (GBSx1174) was identified in *S. agalactiae* <SEQ ID 3401> which encodes the amino acid sequence <SEQ ID 3402>. Analysis of this protein sequence reveals the following:

10 Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.59 Transmembrane 3 - 19 (3 - 19)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.1235 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10251> which encodes amino acid sequence <SEQ ID 10252> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07024 GB:AF001518 unknown: conserved protein [Bacillus halodurens]
 Identities = 86/275 (31%), Positives = 143/275 (51%), Gaps = 6/275 (2%)
 25 Query: 17 MSYPYKANHSIESITLKVNDLENLVNFYSIDIIIGLTVIDKSSTRALLGVNQKIPLILEKT 76
 M + + N ++ + +KV+DL + FY +IIG V++S A L N + PL++E+
 Sbjct: 1 MSFHRQNTFVDLVNFKVSDLSRALTFYQETIGFQVLEERSRSATLTANGRTPLLVIEQP 60
 30 Query: 77 E--LEKHSSTYGLYHTAILVPDEYHLSALNHLLSQHIFLEGAGDHYSNAIYLSDPDGN 133
 + ++ T GLYH A+L+P I LNHL PL+G +DH S AIY +DP+GN
 Sbjct: 61 DPVIAQPRTTGLYHFALLLPBRADLGRFLNHLQSGYPLQASDHLSRAIYFADPDGN 120
 Query: 134 GIEIYNDKDISMWDIRESGQIIGITERLDIDNLDLSLVNPNYKLSKTSIGHIHLSVK 193
 G+E+Y D+ S WD +G++ TR + +NLL + P L +T +GHIHL V
 35 Sbjct: 121 GVEVYADRPSSWD-WSNGEVKMTPEITHAVNLAEKDEFWT-ALPPETILGHIHLVA 178
 Query: 194 DAKISSKLYQNVPFLDEKFAIPT-ASWIASGNYHHHLAPNWAGPNLSKNQSDRFGISLI 252
 + + Y G + + A +I++GNYHHR+ N WG E G+
 40 Sbjct: 179 NLFEPATFYIEGLGPNVVARLGNQALFISTGNYHHHIGLNTWNGVAPTTPSHSVGLKWP 238
 Query: 253 TIAYNDNLFPRDSLKAQYQLTFLEKQDHYIIE 287
 ++ Y + + ++ + + K ++I+
 Sbjct: 239 SLTYFSEVRKATVNRLETIGFQVERKHGGRWVID 273

- 45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3403> which encodes the amino acid sequence <SEQ ID 3404>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0936 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1231-

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/282 (50%), Positives = 194/282 (68%)

```

5  Query: 17 MSYPYKANHSTISITLKVNDELNLVNFYSDIIGLVIDKSSTRALLGVNQKIPILILEKT 76
   M VLY + S + ++L V DL + FY+ IIGL V+ +T L + K ++ L +T
   Sbjct: 1 MIYPYNSTISLGTIVSLVNLDAKMTTFYTSIIGLQVLSQDTTSRQITTDGKTVILELRQT 60

10 Query: 77 ELEKHSITGLYHTAILVPDRHYHSLANHLHLSQHPIEGGADHGYSNAIYLSDPGNGIE 136
   L YGLYHTA WPD + L L LNH L++ I LEG ADHG+S AIYLSDPGNGIE
   Sbjct: 61 PLPGDKAYGLYHTAFLVPRHSLGLVLAHFLTRISLEGGADEGHSEAIYLSDPGNGIE 120

15 Query: 137 IYNDKDISMWDIRESSQIGITERLDYDNLDSLVNVPNNYKLSKTSIGHIHLVSKDAK 196
   IY+DK + WDIR++GOIIG+TE D ++L+ L ++P ++ L++ T I H+HLSVK+A
   Sbjct: 121 IYHDKAVEHWDIRDNDQIIGVTEPTDTKSLIEQLTDIPKHFLAQDTRIRHIVHLSVKNAL 180

20 Query: 197 ISSKLYQNVFGLDEKFAIPTASWIASGNVYHHILAFNHWAGNLSKNQEDRPGISLTIAY 256
   SS LYQ VF L +K IP+ASWIASGNV+HHLAFN+W+ P L K+QE PG++ LTI
   Sbjct: 181 ASSLLYQKVFGLDGRMTIPASWIASGNVYHHILAFNHWASPYLKKIIEGAPGLAFLTIHI 240

25 Query: 257 NDDNLFPSDLKQQAQLQTLFLEKQDMYIIIEPDFGIRIKVVL 298
   LF +LKA+L+ L L++ + ED +GIR+ V+L
   Sbjct: 241 ETPLILFSATLKQARLKLAILQEDSSFTTEDEGIRVNVIL 282

```

SEQ ID 3402 (GBS429) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 7; MW 34.2kDa).

GBS429-His was purified as shown in Figure 214, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1100

30 A DNA sequence (GBSx1175) was identified in *S. agalactiae* <SEQ ID 3405> which encodes the amino acid sequence <SEQ ID 3406>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>> Seems to have no N-terminal signal sequence

```

```

35 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2362 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

40 A related GBS nucleic acid sequence <SEQ ID 10249> which encodes amino acid sequence <SEQ ID 10250> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA021682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]
Identities = 85/261 (34%), Positives = 151/261 (57%), Gaps = 4/261 (1%)

```

```

45 Query: 10 MVRLLIFSDIDGTLINSNFVTPKTRQGIQIVAGQATFVPISARMPEALTIPIMBOIGIDS 69
   M + +FSD +GTL+ S ++P+T IK++ A G FVPISAR P I P +Q+ ++
   Sbjct: 2 MYKAVFSDNGFTLLTTSQHTISPRTVVVKRIITANGIPFVPISARSPILGILFYNKQLETN 61

50 Query: 70 YIISTNGALIQMQQKQTIASHIMDGQVALQVCSYVSIGHYKIAMVNVYRHWYSCKENE 129
   ++++GALL + + I S ++ + L++ + ++H + N Y ++ ++ D EN+
   Sbjct: 62 VLVAFSGALILNCRLEPIYSVOIEPKDILINTVLAISH-PLLOVNYTTNNDCHARDVENK 120

Query: 130 WVQKEEIVGLQSKNSLMETEKQDRIHKLLMGEPSSIMGELENTYKAQYPHLSIAQAP 189
   WV E + ++ + HK+ ++GE +E+E LK ++PHLSI +S
55 Sbjct: 121 WVYERSVTKIEIHPFDEAVTRSP---HKIQIIGAEETIEIEVLLKCKPHLSICTRSHA 177

```

-1232-

Query: 190 YFIRIMAPGIEKGKSAKTLADYLDISLADSIAPGONYNIDANLEIVGKGFVGMNAPKDLQ 249
 F+E+M KG + + L DY + + IAPGON+NDL+LE VG G MGNAP +++
 Sbjct: 178 NPLFVMEKSAATKSSAVRPLDYFGVQTNIEVIAPGONFNLDMLHEVGLGVAMSNAPNEIK 237

Query: 250 ERIGNVTQDNDNDGIYYALVE 270
 + VT N+ DG+ L E
 Sbjct: 238 QANVVVATNNEDGLALLILEE 258

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1101

A DNA sequence (GBSx1176) was identified in *S. agalactiae* <SEQ ID 3409> which encodes the amino acid sequence <SEQ ID 3410>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 19
 >>> May be a lipoprotein
- Final Results -----
- 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:AG07223 GB:AE004801 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 103/283 (36%), Positives = 165/283 (57%), Gaps = 1/283 (0%)
- Query: 33 KHIGILQVVEHPSLTATRKGFIKELAKBOTKDKNLIKIEYVQAQDQSNIQSISEKLIKD 92
 K + + VEHP+L A R G + L + GY+DGEN+K +Y++AQG+ I+ K I D
 Sbjct: 31 KSVAVTAIVEHPALDAARDGVKEALQFAGYEDGKNLKWQYQSAQGVGTAAQIARKFIGD 90
- 30 Query: 93 NK-LVLGIATPAQSLTIVSTETILPTAVTDPVSAELVEMKPKPEGLATOTSDMSPIK 151
 +++GIATP+AQ+L + PI+P+ VTDPV A L E + TG SDM + K
 Sbjct: 91 KPDVIVGIATPSAQALVAATKSIPIVFSTVTDVPGANLTPSWEASGTNVTGVSMDMLALDK 150
- 35 Query: 152 QVSLLRKVMKPKVRVGMVMTTSSERNSEVOVKAKKIPORAGIKTSVKGISSTNDVODAK 211
 Q+ L+KVP+P KR+G++Y E NS V VK+ K++ + G+ + DV A+
 Sbjct: 151 QIELIKKVVPGAKRIQAVNVPGEANSVVVVKELKELLPMGLSLVEASAPRSVDVSSAAR 210
- 40 Query: 212 SLNKSITVIVFPTDNIASSVTLLGNLSKELKVPVVGGSADMVPSGLLPSYGADYEAALGR 271
 SL+ K + I+ TDN + S+ L + + K+P++ D V G + + G +Y+ +G+
 Sbjct: 211 SLVGKVDIAITYTNDNNVGSAYEALVKVGNDAKIPLIASOTDSVKRGKIAALGINKYKMGK 270
- Query: 272 QTARQAVKILKGGKDVAKVPSEYFQNLKVVNVNEMAKRLGIDVS 314
 QT R V+ILKG+ ++ E NL++ VN A++ G+ +S
 45 Sbjct: 271 QTGMVVRILKGGKGEIKPETSNDLQLFVNPGRAGQKQGVILS 313

There is also homology to SEQ ID 2712.

SEQ ID 3410 (GBS188) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 2; MW 36.6kDa).

- 50 The GBS188-His fusion product was purified (Figure 204, lane 6) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 247), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

-1233-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1102

A DNA sequence (GBSx1177) was identified in *S. agalactiae* <SEQ ID 3411> which encodes the amino acid sequence <SEQ ID 3412>. This protein is predicted to be probable permease of ABC transporter (rbsC). Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
10  INTEGRAL    Likelihood = -16.13    Transmembrane    132 - 148 ( 124 - 160)
    INTEGRAL    Likelihood = -6.42    Transmembrane    241 - 257 ( 238 - 258)
    INTEGRAL    Likelihood = -6.32    Transmembrane    264 - 280 ( 260 - 284)
    INTEGRAL    Likelihood = -6.00    Transmembrane    213 - 229 ( 207 - 235)
    INTEGRAL    Likelihood = -4.67    Transmembrane    58 - 74 ( 57 - 75)
15  INTEGRAL    Likelihood = -1.38    Transmembrane    36 - 52 ( 36 - 52)
    INTEGRAL    Likelihood = -0.85    Transmembrane    90 - 106 ( 87 - 106)

----- Final Results -----
          bacterial membrane --- Certainty=0.7453(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA07224 GB:AB004801 probable permease of ABC transporter
[Pseudomonas aeruginosa]
25  Identities = 114/285 (40%), Positives = 175/285 (61%), Gaps = 3/285 (1%)

Query: 5  ILSGISQGLWSINAIQGVFITRILDIADLSAREGAPFMGAALVCIVNDINPIVATIA 64
      + + GL+ S++ A+GVFI+FR+L DL+ +G+FF+G AVCA I +P AT+A
30  Sbjct: 6  LFGALIEGLIFSLVALGVFISFRILRFDDITVDGSFPLGAVCMTIALGMDPYSATLAA 65

Query: 65  MLGGMLAGLVSGFLHTMKIPALLTGITITLTQLYSINLLVLRSSVGFALNIVLTMVTR 124
      G LAGL +G L+ K+KI LL I+ + LYSINL ++G+ NV + TL T++
35  Sbjct: 66  TAAGALAGLATGLINVLKIMDLIASILMMIALYSINLRIMGKPSVPLIAEPTLFTLLQP 125

Query: 125  LGLNKL SAVLLIGIVCVGLVILILYLFNLQTGLALRATGNEAMGQANSIKVDRMKMIG 184
      L+ + V L+L F TQ GLR+RATG N M+A + M+LG
40  Sbjct: 126  EWLSIVYFRPLLLVFIVIAAKILLDDWFFTTQGLAIRATGNSPRMARAGQVNTGMILLG 185

Query: 185  YNIGNGLIALSGALLAQNGYADIANMGVGTIVIGLASIILAEVMIKYLPGLKRLMSIVLG 244
      I N L+AL+GAL AQ G AD++MG+GTVIGLA++I+ E ++ L +++IG
45  Sbjct: 186  MAISGALVALAGALFACTGGADISMGIGTIVIGLAIVTVGESILFRRLLIATLAVILG 245

Query: 245  SVLYRMIIVFILITD---IDAQMIKLVSAILLALILYVPELRAKL 286
      +++YR I L+D + AQ + LV+A+L+ + L P ++ +L
Sbjct: 246  ATYTRFFPTALAINSDFIGLQAQDINLVAVLTVVAVLIVPMKKRL 290

```

There is also homology to SEQ ID 2716.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1103

A DNA sequence (GBSx1178) was identified in *S. agalactiae* <SEQ ID 3413> which encodes the amino acid sequence <SEQ ID 3414>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence
55

```

-1234-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3798(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF06640 GB:AF162694 ABC transporter [Enterococcus gallinarum]
 Identities = 171/264 (64%), Positives = 213/264 (79%), Gaps = 1/264 (0%)

Query: 3 LLELVNHLKTFEKGTVNENHVLRLGLDLTIEDGDFISVIGNGAGKSTLLNACIAGLIPDQ 62
 +L + +LH+TFEKGTVNENHVLRLGLDLTIEDGDFISVIGNGAGKSTLLNACIAGLIPDQ 62
 Sbjct: 5 VLTISDLHCTFEKGTVNENHVLRLGLDLTIEDGDFISVIGNGAGKSTLLNACIAGLIPDQ 64

Query: 63 GAITLDNQSIITKDSVEKRSKDISRVQDPRMGATNLTIEENMAIAHKGKRNHPIRQSV 122
 G I L + + IT + SV +RSK+ISRVQDPRMGATNLTIEENMAIAHKGKRNHPIRQSV 122
 Sbjct: 65 GKIVLGDKEITRHSVTRRSKEISRVQDPRMGATNLTIEENMAIAHKGKRNHPIRQSV 123

Query: 123 TDDDRQLFKKSLSLGLGLLENRMKTDAAFLSGGQQRQALTMATIVRPKLLLEDHTAL 182
 R FK + L + L LLENR + T + LSGGQQRQ + TL MATL + PKL + LLEDHTAL 182
 Sbjct: 124 KGKRAPFFKELKRLNGLLENRLTTEIGLLSGGQQRQALTMATILMATLQPKLLLEDHTAL 183

Query: 183 DPKTSIDWMLTQKVIEBQR/LTALMITHMEHAIAYGKRLVMLYHGKIVVDVKGKAKRLN 242
 DPKTS VM LT + + I + BQ + LTA M + TH + ME AI YGNRL + ML + GKIVVD + GE K + L 242
 Sbjct: 184 DPKTSMTVMALTDQLIEBQR/LTAPMVTHMEDAIRYGNRLIMLHGKIVVDITGEEKQSL 243

Query: 243 TVAEMLFPHQNSGQQLIDDAIHL 266
 TV + LM LPH + NSG + L DD L + L 266
 Sbjct: 244 TVFDLMALFPHQNSGTLEKDDQLL 267

There is also homology to SEQ ID 2720:

Identities = 116/249 (46%), Positives = 166/249 (66%), Gaps = 1/249 (0%)

Query: 3 LLELVNHLKTFEKGTVNENHVLRLGLDLTIEDGDFISVIGNGAGKSTLLNACIAGLIPDQ 62
 +EL+N + G + +L + LTI + DF+++GGNGAGKSTL N IAG + + +
 Sbjct: 4 IIEELINATVDVDMGFEDAKTILDNVLTITIEHDFITLGGNGAGKSTLNFVNIAGTILSLTR 63

Query: 63 GAITLDNQSIITKDSVEKRSKDISRVQDPRMGATNLTIEENMAIAHKGKRNHPIRQSV 122
 G I + Q + T EKR + +SRVQD +MGA +T+ EN+ IA +EG KR + + +
 Sbjct: 64 QGIRILGQDVTHWPAKRALISRVQDSKMGATAPMTVAENLLTARQGGKRSGLASKI 123

Query: 123 TDDDRQLFKKSLSLGLGLLENRMKTDAAFLSGGQQRQALTMATIVRPKLLLEDHTAL 182
 T+ F+ + + G GLR + +T A LSGGQQRQAL L MATL + P LLLLEDHTAL 182
 Sbjct: 124 TEHLAS -FEDLVKRTNGLEKHLETPAGLLSGGQQRQALSLMATLTKPKALLLEDHTAL 182

Query: 183 DPKTSIDWMLTQKVIEBQR/LTALMITHMEHAIAYGKRLVMLYHGKIVVDVKGKAKRLN 242
 DPKTS +M+LT + + + LITLNIH+ME A+ YGNRL+ + + G I + D+ K L 242
 Sbjct: 183 DPKTSQSLMQLTDFPTKDG/LTALMITHMEHEDALTYGNRLIVMGKGNIIKDLNMGKQQL 242

Query: 243 TVAEMLF 251
 T+ + +LF 251
 Sbjct: 243 TITDYQLF 251

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1104

A DNA sequence (GBSx1179) was identified in *S.galactiae* <SEQ ID 3415> which encodes the amino acid sequence <SEQ ID 3416>. This protein is predicted to be mannose-specific phosphotransferase system component IIA. Analysis of this protein sequence reveals the following:

Possible site: 54

-1235-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3527 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD46485 GB:AF130465 mannose-specific phosphotransferase system
component IIB [Streptococcus salivarius]
Identities = 287/336 (85%), Positives = 306/336 (90%), Gaps = 6/336 (1%)

Query: 1 MGIGIIASHGKFAEGIHQSGSMIPGQEKQVQVTFMPNEGPDLYGHFNNAIQFDADD 60
15 MGIGIIASHGKFAEGIHQSGSMIPGQEKQVQVTFMPNEGPDLYGHFNNAIQFDADD 60
Sbjct: 1 MGIGIIASHGKFAEGIHQSGSMIPGQEKQVQVTFMPNEGPDLYGHFNNAIQFDADD 60

Query: 61 EVLVLADLWSGSPFNQASRVNGENPDRMAITGLNLPLIQAQYTERMMDNAGVQVAA 120
E+LVLADLWSGSPFNQASRV GENP+RK+AITGLNLPLIQAQYTERMMDNAGVQVAA 120
15 Sbjct: 61 EILVLADLWSGSPFNQASRVNGENPDRMAITGLNLPLIQAQYTERMMDNAGVQVAA 120

Query: 121 NIIKESKGIKALPEELNPVVEATPVAGVADVPAEVKQSGSIPGTVIGDGKLNINLAR 180
NIIK+K GIKALPEELNP E T A V A P G+IPGTVIGDGKLNINLAR 180
20 Sbjct: 121 NIIKESKGIKALPEELNPVEET+AAPVEAAP+-----QGAIPGTVIGDGKLNINLAR 174

Query: 181 IDTRLHLHGQVATAMTPASKANRIIVASDEVSKDELRLKQIKQAAPGVKANVVPISKLIE 240
+DTRLHLHGQVAT WTPASKA+RIIVASD+V+KDELRLK+LIKQAAP GVKANVVPI KII+
25 Sbjct: 175 IDTRLHLHGQVATWTPASKADRIIVASDEVSKDELRLKQIKQAAPGVKANVVPISKLIE 234

Query: 241 VADCPFGNTRALILFETVQDALRAIEGQVPELNVGSMASHGTGKTVNVLGMDKDDV 300
+KCPFGNT ALILFETVQDALRAIEGQV I ELNVGSMASHGTGKTVNVLGMDKDDV
30 Sbjct: 235 ASKCPFGNTRALILFETVQDALRAIEGQVPELNVGSMASHGTGKTVNVLGMDKDDV 294

Query: 301 AAFKRLDLGVFSDVRKVPND+KK+LPLIL KANV+
A FPKRLDLGV FDKRVKVPND+KK+LPLIL KANV+
35 Sbjct: 295 ACFKRLDLGVFSDVRKVPND+KK+LPLIL KANV+ 330

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3417> which encodes the amino acid sequence <SEQ ID 3418>. Analysis of this protein sequence reveals the following:

Possible site: 14

40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3533 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/336 (85%), Positives = 308/336 (90%), Gaps = 6/336 (1%)

50 Query: 1 MGIGIIASHGKFAEGIHQSGSMIPGQEKQVQVTFMPNEGPDLYGHFNNAIQFDADD 60
MGIGIIASHGKFAEGIHQSGSMIPGQEKQVQVTFMPNEGPDLYGHFNNAIQFDADD 60
Sbjct: 1 MGIGIIASHGKFAEGIHQSGSMIPGQEKQVQVTFMPNEGPDLYGHFNNAIQFDADD 60

Query: 61 EVLVLADLWSGSPFNQASRVNGENPDRMAITGLNLPLIQAQYTERMMDNAGVQVAA 120
E+LVLADLWSGSPFNQASRV GENP+RKMAITGLNLPLIQAQYTER+MDA AGVQVAA 120
55 Sbjct: 61 EILVLADLWSGSPFNQASRVNGENPDRMAITGLNLPLIQAQYTERMDNAGVQVAA 120

Query: 121 NIIKESKGIKALPEELNPVVEATPVAGVADVPAEVKQSGSIPGTVIGDGKLNINLAR 180
NIIKESK+GIKALPE+LNPV E V + G+IP GTVIGDGKLNINLAR 180
60 Sbjct: 121 NIIKESKGIKALPEELNPVEET+AAPVEAAP+-----QGAIPGTVIGDGKLNINLAR 174

Query: 181 IDTRLHLHGQVATAMTPASKANRIIVASDEVSKDELRLKQIKQAAPGVKANVVPISKLIE 240
+DTRLHLHGQVATAMTPASKA+RIIVASDEV+D+LRKQIKQAAPGVKANVVPISKLIE 240
Sbjct: 175 VTRLHLHGQVATAMTPASKADRIIVASDEVSKDELRLKQIKQAAPGVKANVVPISKLIE 234

-1236-

Query: 241 VAKDPRFGNTALILFETVCOALRAIEGGVKEIFELAVGSGMAHSTGKTMVNRVLSMDKDDV 300
 +KDFRFGNT ALILF+T QDALRA+EGGVKEI ELAVGSGMAHSTGKTMVNRVLSMDK+DV
 Sbjct: 235 ASKDFRFGNTALILFPQTQDALRAVGGVKEINELAVGSGMAHSTGKTMVNRVLSMDKEDV 294

Query: 301 AAFELRLDLGVSTFVRRKVPNDACKNLFDLINKANVK 336
 A FEKRLDLGV+PDRKVPND+KNLF+LI K N+K
 Sbjct: 295 ATFEKRLDLGVTFVRRKVPNDACKNLFELIKTNK 330

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1105

A DNA sequence (GBSx1180) was identified in *S.agalactiae* <SEQ ID 3419> which encodes the amino acid sequence <SEQ ID 3420>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 52
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 20 bacterial cytoplasm --- Certainty=0.3873 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:BA06625 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 89/267 (33%), Positives = 139/267 (51%), Gaps = 3/267 (1%)
- Query: 3 KKI+AVDLDGTLHRRNTISDYTADTLRKVQAOCHKVLIITGRPRMALAHYLRDLKTF 62
 +IA+DLDCLE+N TIS T T+K+ QH V+I+TCRFYR++ +Y L L T
 Sbjct: 4 RHLIALDGLTALTNKNTISMKTKQTIQAREAGHIVISTGRPRASIQYQELQDLTA 63
- 30 Query: 63 MINFNGALTHIPEKKWAFERSATIDKLLLETNLSDAIQADFLASEYRNFIYTMNRD 122
 +NFNGA H P+ ++ ++ +A I E ++Y+ D
 Sbjct: 64 IVNFNGAPVHHPDSSPOTYHHIPLSLSTARQVETCEAPDVSNIMVEIDDTYLR--YD 121
- 35 Query: 123 KINPOLFGVNEITDMALDVTKITRNPNALLQTRHKDKYELAKELRQHPNHELEVDWG 182
 ++ Q F + + + K+ + P +I+ + RL L ++ +WG
 Sbjct: 122 ELFIQTPTTEGQGFVEHGNLLKLRDDPTCVLIHPKDDHVELRSLDGAHAVIDQRTWG 181
- 40 Query: 183 GPLNILESPFGVKNKAYAKHLLSLNLQENLIARFDEHNDTEMLAFARTGVAMKIANP 242
 P N+E G+KKA LK + + +E +IARFDE ND EM+ A G AN KA
 Sbjct: 182 AFWNVIEIVKAGMKNKGLKRIADYQVPKERIARFDEHNDFMIEVAKGVMANANID 241
- Query: 243 TLLPYADQIQWINEEDVARTLEKLL 269
 L A+ I +NE+DG+A LE+ L
 45 Sbjct: 242 PLKALAN-DITLSNEEDQAVYLEAL 267

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3421> which encodes the amino acid sequence <SEQ ID 3422>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 50
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 55 bacterial cytoplasm --- Certainty=0.4380 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 188/270 (69%), Positives = 224/270 (82%)

-1237-

Query: 1 MTKKIIAVDLGTLHNNNTISDYTAUTLRKVQACGHKVIITGRPYRMALAHYLRDLK 60
 MTKK-IA+DLGTLH++NTIS YT T++ VQ +GH VII+TGRPYRMAL +YL+L+LK
 Sbjct: 1 MTKKLIADLDGTLHNDHTISTYQTKIKAVQKGEHVIIITGRPYRMALGYLQLNLK 60

5 Query: 61 TPMINFGALTHIPEKKWAFRGRATIDKKLLFTINLEDAIQADFIASEYRNPYITMDN 120
 TP+I FNGALTH+PE+KWA+E +T+DK LL L D Q DFIASEYRNP YITM N
 Sbjct: 61 TPITTFNGALTHMDEQKWAYSHNVLDKGYLLRLKQYQDDPQMDFIASEYRNPYITM 120

10 Query: 121 RDKINPOLFGVNEKITDKMALDVKITRNPALMOTRHKDKYELAKELRQHFNHELEVD 180
 + I+POLFGV+EIT MAL++TKITRNPALMOT H+DKY LAK +R F E+E+DS
 Sbjct: 121 PRSDIPQFGVDREITQDMALEITKITRNPALMOTRHKDYALAKMRACFKDEIRDS 180

Query: 181 WGGPLNILEFSPKGVNKAVALKHLKSLNLSQENLIARQDEHNDTEMLFAHTGYAMINA 240
 WGGPLNILE S K VIKAYAL +LL N+ ++LIARQDEHNDTEMLFA TOYAMINA
 Sbjct: 181 WGGPLNILESSKGVNKAVALNVLGIYNMCKDLIARQDEHNDTEMLFACTGYAMINA 240

15 Query: 241 NPTLLPYADQCIQNTNEEDGVAKLEKLL 270
 +P LLLPYADQ++ ++NEEDGVAK LE+L L
 Sbjct: 241 SPVLLPYADQQLNFSNEEDGVAKLEELFL 270

20

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1106

A DNA sequence (GBSx1181) was identified in *S.agalactiae* <SEQ ID 3423> which encodes the amino acid sequence <SEQ ID 3424>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -7.38	Transmembrane	96 - 112 (90 - 119)
INTEGRAL	Likelihood = -6.58	Transmembrane	28 - 44 (27 - 47)
INTEGRAL	Likelihood = -6.26	Transmembrane	176 - 192 (174 - 193)
INTEGRAL	Likelihood = -5.26	Transmembrane	127 - 143 (126 - 144)
INTEGRAL	Likelihood = -1.59	Transmembrane	4 - 20 (3 - 20)
INTEGRAL	Likelihood = -0.22	Transmembrane	60 - 76 (59 - 78)

30

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1107

45 A DNA sequence (GBSx1182) was identified in *S.agalactiae* <SEQ ID 3425> which encodes the amino acid sequence <SEQ ID 3426>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2025(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

-1238-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1108

- 5 A DNA sequence (GBSx1183) was identified in *S.agalactiae* <SEQ ID 3427> which encodes the amino acid sequence <SEQ ID 3428>. This protein is predicted to be an integral membrane protein. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.
 10 INTEGRAL Likelihood = -5.41 Transmembrane 180 - 196 (179 - 199)
 INTEGRAL Likelihood = -5.31 Transmembrane 96 - 112 (94 - 114)
 INTEGRAL Likelihood = -2.18 Transmembrane 129 - 145 (129 - 145)
 INTEGRAL Likelihood = -1.33 Transmembrane 37 - 53 (37 - 53)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.3166 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 8729> which encodes amino acid sequence <SEQ ID 8730> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7
 MG: Discrim Score: 5.85
 GVH: Signal Score (-7.5): -2.39
 25 Possible site: 18
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 4 value: -5.41 threshold: 0.0
 INTEGRAL Likelihood = -5.41 Transmembrane 176 - 192 (175 - 195)
 INTEGRAL Likelihood = -5.31 Transmembrane 92 - 108 (90 - 110)
 30 INTEGRAL Likelihood = -2.18 Transmembrane 129 - 145 (129 - 145)
 PERIPHERAL Likelihood = 0.05 57
 modified ALOM score: 1.58
 *** Reasoning Step: 3
 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.3166 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
 40

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC5028 GB: AE001188 conserved hypothetical integral membrane
 protein [Treponema pallidum]
 45 Identities = 54/190 (28%), Positives = 93/190 (49%), Gaps = 14/190 (7%)
 Query: 14 LFFTVISPGIKYYHLQG--PMLIWNMTALIALDFAYLTSL--PKKKILIGLFLAWFF 69
 + F + + + S P G + L + W N + L A I + + + F + + L W F
 Sbjct: 3 VFCILLISFGRCVAADNPLSFLVWNLVLAFLPWLISAILHVRFRFAVRSVQLFLMLLWLLG 62
 50 Query: 70 YPNTFYMLTDIIHMHFVGVDVLYNKINILYILYVSSILFGFLSGISSPVMKRFKRISNI 129
 + P N Y + L T D I I H + L + I L + + + F + S S + + R F I
 Sbjct: 63 FPNAPYIILTDIILHGKKGKSPLLYYDILILLAYSFTGLPYAFVSLHLSLILARDPHIKRP 122
 55 Query: 130 FLRWGIIGTVSL-VSSFGIHIGRYARINSDWILTKPQVVINELLAVFSR-----DSFHF 183
 F I I + L + + FGI + + GR + R NSWDI + + + + + + R D + + F +
 Sbjct: 123 F ---- I I S V F E L C A P G I Y L G R F L R W N S D W I L H G R T I L S D I G I R V I R P V F V T D W M F V 178
 Query: 184 LGFTPLQVLC 193

F + VLC

Sbjct: 179 FFFGTMLVLC 188

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1109

A DNA sequence (GBSx1184) was identified in *S.agalactiae* <SEQ ID 3429> which encodes the amino acid sequence <SEQ ID 3430>. Analysis of this protein sequence reveals the following:

10 Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.79 Transmembrane 171 - 187 (166 - 191)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.3718 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1110

A DNA sequence (GBSx1185) was identified in *S.agalactiae* <SEQ ID 3431> which encodes the amino acid sequence <SEQ ID 3432>. Analysis of this protein sequence reveals the following:

25 Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 30 INTEGRAL Likelihood = -10.46 Transmembrane 193 - 209 (191 - 214)
 INTEGRAL Likelihood = -10.30 Transmembrane 99 - 115 (96 - 119)
 INTEGRAL Likelihood = -8.17 Transmembrane 454 - 470 (451 - 472)
 INTEGRAL Likelihood = -6.64 Transmembrane 216 - 232 (212 - 236)
 INTEGRAL Likelihood = -6.37 Transmembrane 49 - 65 (43 - 68)
 INTEGRAL Likelihood = -4.88 Transmembrane 362 - 378 (357 - 383)
 35 INTEGRAL Likelihood = -3.61 Transmembrane 385 - 401 (385 - 402)
 INTEGRAL Likelihood = -2.76 Transmembrane 275 - 291 (275 - 291)
 INTEGRAL Likelihood = -1.70 Transmembrane 18 - 34 (18 - 34)
 ----- Final Results -----
 40 bacterial membrane --- Certainty=0.5182 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAF95422 GB:AR004299 conserved hypothetical protein [Vibrio cholerae]
 Identities = 193/471 (40%), Positives = 286/471 (59%), Gaps = 42/471 (8%)
 Query: 1 MEKFFKLEKHGTTIRTELTAAGLTFFAMSYLLFVNPAILSQTGMPAQGVFLATIGAVVA 60
 +EK PKL +E+GT +RTEI AG+TTF M+YI+FNPAILS GM VF+AT + A +
 Sbjct: 2 LEKLEFKLSEYGTNVRTELTAAGVTTFLTMAYIIFVNPAILSDAGMDRGAVPVATCLAAAG 61
 50 Query: 61 TSVGAFYANLPYAQAQPGMGLNAFFTYTVVFAIGYTWQKALAMVFICGLISLIYTLTKVRK 120
 +M F AN P AQAQPGMGLNAFFTY VV +G+TWQ ALA VF G++ ++++L K+R+
 Sbjct: 62 CFIMGFLIANYPIAQAQPGMGLNAFFTYGVVLMGHTWQVLAALAVFCSGVLFIILSLFKIR 121

-1240-

Query: 121 MIISIPITLKSAITAGIGTFLAYVGIKNAGLKFSPIDPGTYDVVKGGAARKLATITANS 180
 II SIP L+ + I+AGIG FLA+ + +KNG + +P T +V GA L +
 5 Sbjet: 122 WIINSIPHSRLTGISAGIGLFLAFLAKNAGIV -VDNPAT -LWSIGATSLHAV - 173

Query: 181 SATPGLVSPDNPAILLISLIGLSITIPFVKIRGQIILSILMTTLGLTILMGVVKLADINW 240
 L+ +G +TI + +G+G + + +IL T LG+ G V+ I
 10 Sbjet: 174 -----LAAVGFPLTIGLVYRGVKGAVMIALLAVTALSLVPGDVWGQIMS 218

Query: 241 EATNLASAFRDLEQVFGVALGKGLISLSPNPKRLPSVLMAILARSLTDIFDTIGLIGT 300
 + + + F Q+ A+ E G+IS+ + AF D+FDT GIL+G
 10 Sbjet: 219 TPPSIAPTF -MQLDPSAVFEIGMISV - - - - -VFALFVLDLFDTAATLNGV 262

Query: 301 GEKVGILATTGDNHRSKSLDKALYSDLIGTFCGACGTSNVITTVESAGIGAGGTGLT 360
 K G+ G + L+AL +D T+ GA+ GTSN T+Y+ES +G+ GRTGLT
 15 Sbjet: 263 ATKAGLIEKDG - - - KIPRLNALLADSTATSVGALLGTSNITTSYIESVSGVAVGKTGLT 319

Query: 361 ALVVGILFAISSFPSPSIVPSQATAPILVIVGINMLGNLKDTRKDDNSEALPAFFTS 420
 A+VV LF + + FFSPL + +P+ ATA L V I+N+S L I W D+EA P T L
 20 Sbjet: 320 AVVVGILFLALFSPPLAGMIIPAYATAGALFYVAILAMSGSLVSDIKRDLTEAAPTFFVCL 379

Query: 421 FMGFTSYITYGIAAGFLTYTLAKVIGQAKDHRVVMILDLFLINFLISLA 471
 M T+SI GI+ GF+ Y K+ G+ + + +W+ +F+ +I A
 25 Sbjet: 380 NMPLTPSIABGSLGFLIYDAIKLPSGKGRSVLSGVWMAIFVVIKILAA 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3433> which encodes the amino acid sequence <SEQ ID 3434>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

30	INTEGRAL	Likelihood = -11.57	Transmembrane	378 - 394 (370 - 419)
	INTEGRAL	Likelihood = -9.29	Transmembrane	202 - 218 (195 - 221)
	INTEGRAL	Likelihood = -7.64	Transmembrane	48 - 64 (46 - 71)
	INTEGRAL	Likelihood = -7.64	Transmembrane	99 - 115 (97 - 118)
	INTEGRAL	Likelihood = -6.90	Transmembrane	225 - 241 (221 - 245)
35	INTEGRAL	Likelihood = -6.05	Transmembrane	468 - 484 (465 - 485)
	INTEGRAL	Likelihood = -4.35	Transmembrane	399 - 415 (395 - 419)
	INTEGRAL	Likelihood = -3.24	Transmembrane	425 - 441 (425 - 442)
	INTEGRAL	Likelihood = -3.08	Transmembrane	18 - 34 (18 - 34)
	INTEGRAL	Likelihood = -2.28	Transmembrane	442 - 458 (442 - 460)
40	INTEGRAL	Likelihood = -0.00	Transmembrane	282 - 298 (282 - 298)

----- Final Results -----
 bacterial membrane --- Certainty=0.5628 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GF:BA04327 GB:AP001509 unknown conserved protein [Bacillus halodurans]
 Identities = 192/485 (39%), Positives = 276/485 (56%), Gaps = 53/485 (10%)

50 Query: 1 MEKFFKLSGNTVSTETINAGLITTFPMGYILFVNPSILGAGMPSNVAFLATILAAIS 60
 N+ + + F E+GTT E +AGLITF + +YILFVNPS ILG AGM AVF+AT +AAAI
 Sbjet: 1 MDRYFGPKKHGHTYGRSLAGLITFLSMAYILFVNPLILGDAGMDVQAVPMATILAAIG 60

55 Query: 61 TLIMGLFANVYALAPGMLINAFPTTVFALRFSWQELAMVFIQGLNFHIFTVTKFR 120
 TLIMG+ A P ALAPGMLINAFPT Y+VV + WQ AL VF+ G+ I ITV K R+
 Sbjet: 61 TLIMGILAKYPIALAPGMLINAFPTAYFVIGMIGLQALPGVFGIIFILITVFKIR 120

60 Query: 121 SIKAIPUSLQIAVGGGIVGFVAYLGPKNANITFSIAENIWMVNGVEPAKASAKTPAD 180
 II AIP L+ +A GIG+P+ A+ +G KNA I+
 60 Sbjet: 121 VIINAIPAEIKNAAGIGLFLAFLAKNAGIV - - - - - 154

Query: 181 GLLFVDAGVGVPTISFTSGVLLAIGLGLLITLAVIRNPGAILIGIVATTLNGIPL 240
 + + + + LIA PGL+T + +R +G I G+ T +VG+ G
 65 Sbjet: 155 -----SDEATAVSLGHILNGPTLACPLGILTVLFWVRGICQGIYQGMILTAIVGLIS 208

-1241-

5	Query:	241	IVDVSNLNFGLSHIGEAWEITLGTTFLAADF-GLSSLPSDSSRLPLVFWMTIFAFSLSDTFD	299
	Sbjct:	209	I+ + I L TF AF+ ++FS + + F D FD IITYTG-----GGIVSTPSPSLAPTFGQAFNIQMAVDSVQ-----FLIVLTLFLFVDFD	258
10	Query:	300	TIGTFIGTGRRTGIPSQDDENALENSIGFSSKMDRALFADAIOTSIGALVGTSNITTYVZ	359
	Sbjct:	259	T GT G + G F +D++ + +AL AD+ TSIGA++GTS TT Y+Z TAGTLGVANQAG-FIKNK-----LPRAGKALLADSSATSIGALGTSTTTTIAIR	308
15	Query:	420	SKFADALPAFFAFAFFMALCYSISYGLAAAFIFYCLVKVVEGKTDIHPHIGWATFLFIVN	479
	Sbjct:	369	+K A+P+F M L YSI+ GIA F+FY + +V+G+ K+HPI++ F+P+ TKELALPSFLTVVMPLTYSIATGIAFGFLFYPTIMIVKGRGKEVHPIMYALFFVFLAY	428
20	Query:	480	FIILT 484	
	Sbjct:	429	FI L+ FIFLS 433	

An alignment of the GAS and GBS proteins is shown below.

Identities = 258/488 (52%), Positives = 336/488 (67%), Gaps = 17/488 (3%)

25	Query:	1	MEKFFKLKEHGTTIRTEITAGLITTFPMSYILFVNPAILSOTGMPAGVFLATIIIGAVVA	60
	Sbjct:	1	MEKFFKL E+GTT+ TRI AGLITTFPMSYILFVN+IL GMP+ VFLATII A++ MEKFFKLSENGTTVSTEINAGLITTFPMSYILFVNPSILGAQMPSNAVFLATIIAALIS	60
30	Query:	61	TSVMAFYANLPYAQAQPGMGLNAFFTYTVVEALGYTWOEALAMVFICGLSLIITLTKVRK	120
	Sbjct:	61	T +M +AN+PYA APMGLNAFFTYTVVEAL ++MQEALAMVFICGL ++ IT+TK RK TLINGFANVPYALAPGKGLNAFFTYTVVEALRFSWQELAMVFICGLNFITVTYTKRK	120
35	Query:	121	MIIESIPTTLKSAITAGITGFLAYVGKINAGFLKFSIDPGTYDVV-----GKGAAK	171
	Sbjct:	121	II++IP +L+ AI GIG F+AY+G KNA + FSI +V K A SIIKAIPVSLQHAIGGGIGVFVAYLGFNANIITFSISAENIMVWNGVEPAKASAKTFAD	180
40	Query:	172	GLATITANSSATPGLVSPDNPAILLSLIGLSITTFIVKIGRGIIISLITTLGLILMG	231
	Sbjct:	181	GL + AN P + SP + +LL+ GL +T +++ RG II+ I+ TTL+GI +G GLLFVDANGGVPTIISSTDSGVLLAIFGLLLTTLVIRNFRGAILIGIVATTLVGPIAG	240
45	Query:	232	VVKLDADNKEATNLASFRDLKQVFGVALGKGLISLFENPSRLPSVIMAILAFLSLDIF	291
	Sbjct:	241	+V + +N+ +++ ++ +L F A GL SLFS+ SELP V M I AFSL+D F IVDVSNLNFGLSHIGEAWEITLGTTFLAADF--DGLSSLPSDSSRLPLVFWMTIFAFSLSDTF	298
50	Query:	292	DTIGTLIGTGEKVGILATGDN-----HESKSLDKALYSDLIGITFGAIGTSNVTTYV	345
	Sbjct:	299	DTIGT IGTG + GI + +N S +D+AL++D IGT+ GA+ GTSN TTYV DTIGTFIGTGRRTGIPSQDDENALENSIGFSSKMDRALFADAIOTSIGALVGTSNITTYV	358
55	Query:	346	ESAAGIGAGGRTGLTALVVAGLFAISSFTSPLVSIVPSOATAPILVIVGIMMNLNKDIK	405
	Sbjct:	359	ESAAGI GORTGLTA+ A F+S PLV IVP+ ATAP L+IVG+MM+S+ D+ ESAAGIABGGRTGLTAVSTAVCFLLSILLPLVGI VPAATAPALIVGVNVMVSSFLDVN	418
60	Query:	406	WDMSEAIAPAFFTSLPMGFTYSITYGIAAGFLTYTLAKVINGQAKDIHVIMLIDILFIL	465
	Sbjct:	419	W ++A+PAFF + PM YSI+YGLAA F+ Y L KV++G+ KDIH ++W LFI+ WSKFADALPAFFAFAFFMALCYSISYGLAAAFIFYCLVKVVEGKTDIHPHIGWATFLFIV	478
	Query:	466	NFISLAIL 473	
	Sbjct:	479	NFI L IL NFILITIL 486	

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1242-

Example 1111

A DNA sequence (GBSx1186) was identified in *S.agalactiae* <SEQ ID 3435> which encodes the amino acid sequence <SEQ ID 3436>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3221 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA04264 GB:AP001508 unknown conserved protein [Bacillus halodurans]
Identities = 68/147 (46%), Positives = 100/147 (67%), Gaps = 1/147 (0%)

Query: 27 MFYTNNEELIALGQKLGTVLKSQDIVLLTGNLQAGKTTLTGKLAGLQDKNIKSPITYT 86
      M TQ+ E +A QKL L +SD++ L G+LQAGKT+ TKG+A GL IK++KSP+T
Sbjct: 5 MNITQSPETNAFAQLADKLQAGDVITLEGDLQAGKTSFTFKGLALGLGKRVKSPPTT 64

Query: 87 IVREYBSRVPFLYHLDDVYRIGDDPDSIDLDDFLPGQGVTVIEWGELLSDNLIINYLEIVIT 146
      I+REY+GR+FLYR+DYR+ ++ + + D++ G GVT+EW L+ L L I IT
Sbjct: 65 IIREYKRLFLYHMDVYRINEEEDLGDFEYFIKGDGVTVIEWASLIEGRLPFVRLAITIT 124

Query: 147 RSNQG-RQVQLEAYGHRAREIIEAIQD 172
      + + RQ+ AYG R E+++ + D
Sbjct: 125 HAGENRQLSPTAYGERWEEVLKELLD 151

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3437> which encodes the amino acid sequence <SEQ ID 3438>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1202 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 97/142 (68%), Positives = 122/142 (85%)

Query: 27 MFYTNNEELIALGQKLGTVLKSQDIVLLTGNLQAGKTTLTGKLAGLQDKNIKSPITYT 86
      MFY++NE L A G+ L G L T L G D++L+G+LQAGKTTLTGKLAG+ I QNIKSPITYT
Sbjct: 1 MFYSNEYTLKAYGETLGTLSIGDIVLVSQDLQAGKTTLAQKLAGMGISQNIKSPITYT 60

Query: 87 IVREYBSRVPFLYHLDDVYRIGDDPDSIDLDDFLPGQGVTVIEWGELLSDNLIINYLEIVIT 146
      IVREYBSR+FLYHL+YR+GDDPDSIDLDDFLPG GVTIEWGELL + L+ +YL+I IT
Sbjct: 61 IVREYKRLFLYHLDIYRIGDDPDSIDLDDFLPGGVTVIEWGELLQDVLQITIT 120

Query: 147 RSNQGRQVQLEAYGHRAREIIE 168
      ++GRQ+ L A+G R+R+++E
Sbjct: 121 KRDKGRQLDLLAHGERSQLLE 142

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1112

A DNA sequence (GBSx1187) was identified in *S.agalactiae* <SEQ ID 3439> which encodes the amino acid sequence <SEQ ID 3440>. Analysis of this protein sequence reveals the following:

-1243-

Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.1782 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP: AAD35662 GB: AB001732 conserved hypothetical protein [Thermotoga maritima]
 Identities = 56/163 (34%), Positives = 94/163 (57%), Gaps = 1/163 (0%)
- Query: 24 EASREASAILLEFIATVTKETDFLIHTVSNQLSEMTFFIEWTLMKNCILIAKLK 63
 EAS +A I+E+L VT EIDF++ +S +I + ++ ++ +
 15 Sbjct: 18 EASINDARRIVKLEKVTSETDFLITRFDEVIDVSTERNYIRMYRSNPKLMTVGGINRE 77
- Query: 84 VIGLLTIISQSDIEIHVVDLFIAVQKDTWYGIGHILMEBAISWASINDITRRLELSVQ 143
 ++ L+T +HVG++ I+V+K YW GIG ++ ALEWA N R++L V
 20 Sbjct: 78 IVSLLTPTGFGKRKTKVQKIGIGVKKRYNIGIGITRMTISAIEWARRNGPI-RIQLEVL 136
- Query: 144 GRNERAHLVQKFGFEDIGLQTRGKIKRNGEFLDIYRMSKLID 186
 NERAI LN+K GFE+G++ + +R+G F D+ M+L+D
 Sbjct: 137 KNERAISLYRKLGFELGIGKRAVRDDGSDFDVLVWALLLD 179

- 25 There is also homology to SEQ ID 1724.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1113

A DNA sequence (GBSx1188) was identified in *S. agalactiae* <SEQ ID 3441> which encodes the amino acid sequence <SEQ ID 3442>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

- 35 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 40 >GP: CAB15582 GB: Z99122 membrane-bound protein [Bacillus subtilis]
 Identities = 108/324 (33%), Positives = 178/324 (54%), Gaps = 33/324 (10%)
- Query: 5 KKITLMPASAILITTVIALGV--YVASAYNFSTNELSKTFKPKLAKS--KSHATEETKPF 60
 KK TL+ + + + ++ LG Y ++ + + + + + +K K +I+ PF
 45 Sbjct: 8 KKITLLLTILTITLIGLVLTGGYAYYIWHKAASTVASIHESIDSKRDRKEVGSINKDPF 67
- Query: 61 SILLMGVDTGSEHREKSKWNSGNSDMILVTINPKINKITMTSLERDLVILKLGSPKNGQTG 120
 S+L+MGVD + G +D+I +T+NPKTN T M S+ ED K+ G G
 50 Sbjct: 68 SVLIMGVDERDGDK-----GRADTLIYMTVNPKNITMDVSIPTDTYTKIIGK-----G 116
- Query: 121 VEAKINAAAYASGGAEMAITVQDILLINVDYFMQINMQGLVDLVNAGSITVTINKDFPFI 160
 K+N +YA GG +M + TV++ LD+ VDYF+++NM+ D+V+ +GGITV + F F
 55 Sbjct: 117 TMDKINHSYAFGQTQMTVDVTENFLDVFVDYFVKVNMESFRDVIDTLGGITVNSI PAFSP 176
- Query: 181 SIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPGDYGRQKQREVIQKVLKILAL 240
 + G +NG++AL Y+RMR +DP GD+GRQ RQR+VIQ ++ K +
 60 Sbjct: 177 DGYS-----FGKEITLNGKEALYTRMKEDPRGDFGRQDRQROVIGIINGANI 228
- Query: 241 NSISSYKKILSAVSNMOTNRISSKTPINL---LAYKDSLEHISKYQLKGEDATLSDG 296

-1244-

+SI+ + + V NM++TN+ T N+ YK + +HIK ++LKG T +G
 Sbjct: 229 SSTITKFGDMFKVVENNVKTNL-----TFDNMDIQSDYKGARKHIQKHELEKG-TGTFKING 282

Query: 297 GSYQILTCKHLLAVQNRRIKELDK 320
 Y + L + +K+ L+K

Sbjct: 283 IYYVQADESALSDITKELKESLEK 306

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2763> which encodes the amino acid sequence <SEQ ID 2764>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/436 (66%), Positives = 342/436 (78%), Gaps = 22/436 (5%)

Query: 1 MKIWKKITLMSAILLTTVIALGVVYASAYNFSINELSKTFKDFKLAKSKSHAIETKPF 60
 MKI KKI LMF+AI+LTTV+ALGVY+ SAY FST ELKSTFKDF + +KS AI++T+ F
 Sbjct: 1 MKIGKIVLMFTALVLTIVLALGVYLTSAITFTSGELSKTFKDFSTSSNKSDALKQTRAF 60

Query: 61 SILLMGVDTGSEHRKSKWGSNDSMILVTINPKINTMTSLERDVLKLSGPKNNQGT 120
 SILLMGVDTGS R SKW GNSDSMILVT+NPKT KTTMTSLERD L LSGPKNN G
 Sbjct: 61 SILLMGVDTGSSERASKWGSNDSMILVTINPKIKTTMTSLERDVLTLTSGPKNNQNG 120

Query: 121 VEAKLMAAYAAGGACALMTVQCLLDINVDYFMQINMQQLVDLNVAGGITVTNKFDFPI 180
 VEAKLMAAYA+GGA+GA+MTVQCLL+I +D ++QINMQQL+DLNVAGGITVTN+FDPEI
 Sbjct: 121 VEAKLMAAYAAGGACALMTVQCLLDINVDYFMQINMQQLVDLNVAGGITVTNKFDFPI 180

Query: 181 SIAANEPEYKAVVEPOTHRKINGEALVYSRMRYDDPEGDYGRQKQREVIQKVLKCLLAL 240
 SIA NEPEY+A V POTHKINGEALVY+RMRYDDPEGDYGRQKQREVIQKVLKCLLAL
 Sbjct: 181 SIAANEPEYQATVAPOTHRKINGEALVYARMRYDDPEGDYGRQKQREVIQKVLKCLLAL 240

Query: 241 NSISSYKILSAVSNMMQTNIEISSKTIPLNLAYKDSLEHIKSYOLKGEDATLSGGGSYQ 300
 +SISSY+KILSAVS+NMQTNIEISS+TIP+LL Y+D+L IK+YOLKGEDATLSGGGSYQ
 Sbjct: 241 DSISSYKILSAVSNMMQTNIEISSRTIPSLLOYRDLRTIKTYOLKGEDATLSGGGSYQ 300

Query: 301 ILTKHLLAVQNRRIKELDKGRSKTLAKTSAILVEDYDYGTTASNDSSYTSOENYNTT- 359
 I+T HLL +QNRI+ EL + LKT+A +YE+ YG ST S T NNY+++
 Sbjct: 301 IVTSHHLEIQNRIRTELGLHKVNLQKTATVYENLYG-----STKSQTNNYNDSSG 353

Query: 360 ---PYSEAPPYSYG-----NITYSSETNQTHQVNYNSSTPASNYSNNTNQADSSGSV 411
 YS++ SY+ +T S+T+Q + + +TP+S+ S ++ SSGS
 Sbjct: 354 QAPSYSDSHSYANYSGVDTGQASDQSTASSHRAFPSSS-SDALAADSSSSGS- 411

Query: 412 NNHNGAATPNPNTQTQ 427
 G+ P N Q
 Sbjct: 412 ---GSLVPPANINPQ 423

SEQ ID 3442 (GBS54) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 8; MW 48.4kDa).

The GBS54-His fusion product was purified (Figure 98A; see also Figure 194, lane 6) and used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 98B), FACS (Figure 98C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

-1245-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1114

A DNA sequence (GBSx1189) was identified in *S. agalactiae* <SEQ ID 3443> which encodes the amino acid sequence <SEQ ID 3444>. This protein is predicted to be Ves1-L. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -1.44 Transmembrane 3 - 19 ( 3 - 19)
```

```
----- Final Results -----
bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3445> which encodes the amino acid sequence <SEQ ID 3446>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 42/98 (42%), Positives = 64/98 (64%)

Query: 1 MKIGRLIALGLVSLGALVLYQNRKTIKDSYQNTKNETDSAKLKLERIKNDLAIISQKEK 60
      NK+ +IA+GL+S A + Y+ R TIK+ ++ D+A+L L+ IK +L +I + +
Sbjct: 1 MKKVTIAGVLLGSLPTAYKAYQKRCTIKELLSISROAKDRAQDLDNKANKDLIHSQGVK 60

Query: 61 IRLISQELNHHKFOVPFNKDIOPLREEINQRMAYQBEDE 98
      I+ ISQ+L HK++ FN++ Q L EI RMAKYQE R
Sbjct: 61 IQNISQDLAKHWRYFPAQETQAHLETIQNRMAKYQBEDE 98
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1115

A DNA sequence (GBSx1190) was identified in *S. agalactiae* <SEQ ID 3447> which encodes the amino acid sequence <SEQ ID 3448>. This protein is predicted to be Hit-like protein involved in cell-cycle regulation (hit). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2694 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

-1246-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04908 GB:AP001511 Hit-like protein involved in cell-cycle
regulation [Bacillus halodurans]
Identities = 74/137 (54%), Positives = 95/137 (69%), Gaps = 2/137 (1%)

5 Query: 3 NCIFCKIISGEIPSSKVIYEDDEVLAFLDITQTTHGHTLLIPKHHVNVLEMDERTAQITF 62
NCIFCKII+GEIPE+ VYED V AFLDI+Q T GHTL+IPK H RMV E+ E+ A P
Sbjct: 6 NCIFCKIAGEIPSATVYEDDEVVAFLDISQVTKGHTLVIPKHHVNVLESEIASSLF 65

10 Query: 63 ERLPKVARAVQAATKAGKMMIINNNEELAGQTVFHAHVHLVPRFDSGDIKIHVYTHEPD 122
+PK++RA+ A + GMII+NNN E AGQTVF H+HL+PR+ E DG + H
Sbjct: 66 AAVPKISRAINDAPQPIGMIIVNNNGRAAGQTVFPHYHLHLPRYGBGQGYGAVWKDESSQ 125

15 Query: 123 F--BALAKLAKIRKEI 137
+ + L L+ IR+ +
Sbjct: 126 YSGDDLQVLESSIREHL 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3449> which encodes the amino acid sequence <SEQ ID 3450>. Analysis of this protein sequence reveals the following:

20 Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25 bacterial cytoplasm --- Certainty=0.0125 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 97/137 (70%), Positives = 117/137 (84%)

30 Query: 1 MNCIFCKIISGEIPSSKVIYEDDEVLAFLDITQTTHGHTLLIPKHHVNVLEMDERTAQI 60
M+NCIFC II G+IPSSKVIYED++VLAFLDI+QTT GHTL+IPK+HVM+LEM +TA
Sbjct: 1 MNCIFCSIIQGDIPSSKVIYEDQVLAFLDISQVTKGHTLVIPKHHVNVLEMDERTASH 60

35 Query: 61 TFERLPKVARAVQAATKAGKMMIINNNEELAGQTVFHAHVHLVPRFDSGDIKIHVYTHE 120
F R+PK+ARA+Q+AT A MNIINNNE +AGQTVFHAHVHLVPR++E DGI I YTHE
Sbjct: 61 LFAPIPKIARAIQSATGNTAMNIINNNEALAGQTVFHAHVHLVPRYNEEDGISIQVYTHE 120

40 Query: 121 PDPEALAKLAKIRKEI 137
PDP L KLA++I +E+
Sbjct: 121 PDFFVLEKLARQNDQEV 137

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 1116

A DNA sequence (GBSx1191) was identified in *S.agalactiae* <SEQ ID 3451> which encodes the amino acid sequence <SEQ ID 3452>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
50 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55

A related GBS nucleic acid sequence <SEQ ID 10923> which encodes amino acid sequence <SEQ ID 10924> was also identified.

-1247-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3452 (GBS87) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 3; MW 19.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 10; MW 44kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1117

A DNA sequence (GBSx1192) was identified in *S.agalactiae* <SEQ ID 3453> which encodes the amino acid sequence <SEQ ID 3454>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -6.53    Transmembrane    143 - 159 ( 141 - 161)

----- Final Results -----
bacterial membrane --- Certainty=0.3612 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9563> which encodes amino acid sequence <SEQ ID 9564> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP: CAB12844 GB: Z99109 ABC transporter (ATP-binding protein)
[Bacillus subtilis]
Identities = 137/242 (56%), Positives = 181/242 (74%)

Query: 1 MTLKLIENVGGYVNIPLKNISFVNDGELVGLIGLNGAGKSTTINEIIGILRPYQGD I 60
M++L ++++TGGY FVLKN+SF + ++VGLIGLNGAGKSTTI IIG++ P++G I
30 Sbjct: 1 MSLSLVKIDL TGGYTRNPLKNVSTLEPNQIVGLIGLNGAGKSTTIRRHIGLMDPEKSG I 60

Query: 61 TIDGISLEADQELYRKIGFIPETPSLYEELTLREHLETVANAYDIATDEVMAQKLLS 120
++G + D E YR + +IPETP LYEELTL EHLE AMAY ++ + + R LL+
35 Sbjct: 61 ELNGKTFADDPBGYRSQFTYIPETFVLYRELITMEHLELTANAYGLSKETMEKRLPFLK 120

Query: 121 MFRITDKLDWPFMHFSKGMKQKVMICAFVVSPLFVDEPFLGLDPLAISDLINLAE 180
FR+ +L WFP HFSKGMKQKVMICAF+ P+L+I+DEPFLGLDPLAI+ L+ + E
35 Sbjct: 121 EFRMEKRLKWFPAHFSKGMKQKVMICAFLEAPALYI IDEPFLGLDPLAINALLERME 180

Query: 181 KARGKSLMSTHVLDSASKMCOIFVILHKGELTAVAGTLEELRAITPGDSNANINDIYALIT 240
K G S-LMSTH-L +A+R+ CD P+LH GE+RA GTL ELR FG +A L+D+Y+ LT
40 Sbjct: 181 KGGASVLMSTHILATARYCDSPFIILHNGEVRARGTSELREBQPGMKDAALDDLYLELT 240

Query: 241 KE 242
KE
45 Sbjct: 241 KE 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3455> which encodes the amino acid sequence <SEQ ID 3456>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -5.04    Transmembrane    141 - 157 ( 139 - 158)
```

----- Final Results -----

bacterial membrane --- Certainty=0.3017 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB12844 GB:299109 ABC transporter (ATP-binding protein)
[Bacillus subtilis]
Identities = 139/241 (57%), Positives = 189/241 (77%)

Query: 1 MLNIEMLTGGYHNIFVNDVSPVNGELVGLIGLNGAGKSTTINIEIGFLPKYQGSISI 60
+L+K+IRGGY PVL +VSF+++ ++VGLIGLNGAGKSTTII IIG + P+GSI +
Sbjct: 3 LLSVKDLTGGYTRNFVNVSTPLEPNIQVGLIGLNGAGKSTTIRHIGLMDPHKSGIEL 62

Query: 61 DGLTLAENAVAYRQKIGFIPETPSLYEELTSEHINTVAMAYDIDLEVAQKRAQPFLEMP 120
+G T AB+ YR + +IPETP LYEEELT EH+ AMAY + E +R P L+ F
Sbjct: 63 NGKTFADPEGVRSQFTYIPETPVLYKSLTMEHLELTAMAYGLSKETMEKKIPPLKEF 122

Query: 121 RLTDKLENFPNFSKGMKQKVMIIICAFVDPSPSLFLDEPFLGLDPLAISDLIQTLEVEKA 180
R+ +LWFP +PSKGMKQKVMIIICAF+ +P+L+I+DEPFLGLDPLAI+ L+ + K
Sbjct: 123 RMEKRLNFPNFAHPSKGMKQKVMIIICAF+LAEPALYTIIDEPFLGLDPLAINALLERMEAKK 182

Query: 181 KGKSIILMSTHVLDSABRMCDRFVILHKGIVRAQGTADLQRAFGDRSASINDIYALTKED 241
G+LMSH+L +AER CD F+ILH+G+VRA+GTL+L+E FG + A+L+D+YL LTKED
Sbjct: 183 GGASVLMSTHILATAERYCDSFTIILHKGIVRAQGTLSLELREFGMKDAALDDLYLELTKED 243

An alignment of the GAS and GBS proteins is shown below.

Identities = 181/240 (75%), Positives = 208/240 (86%)

Query: 3 MLKIENVVGGYVNIPLVNLNISEFVNDGELVGLIGLNGAGKSTTINIEIGFLPKYQGDITI 62
ML I+N+TGGY NIPVL ++SF V+GELVGLIGLNGAGKSTTINIEIG FLPKYQGS 1+I
Sbjct: 1 MLNIEMLTGGYHNIFVNDVSPVNGELVGLIGLNGAGKSTTINIEIGFLPKYQGSISI 60

Query: 63 DGISLEADQELYRNKIGFIPETPSLYEELTLREHLETVAMAYDIATDEVMAKQKLEMP 122
DG+L + YR+KIGFIPETPSLYEELTL EH+ TVAMAYDI + RAQ LEMP
Sbjct: 61 DGLTLAENAVAYRQKIGFIPETPSLYEELTSEHINTVAMAYDIDLEVAQKRAQPFLEMP 120

Query: 123 RLTDKLENFPNFSKGMKQKVMIIICAFVDPSPSLFLDEPFLGLDPLAISDLIQLAEKA 182
RLTDKL+WFP++PSKGMKQKVMIIICAFV+ PSLFT+DEPFLGLDPLAISDLI L EKA
Sbjct: 121 RLTDKLENFPNFSKGMKQKVMIIICAFVDPSPSLFLDEPFLGLDPLAISDLIQTLEVEKA 180

Query: 183 KGKSIILMSTHVLDSABRMCDRFVILHKGIVRAQGTADLQRAFGDRSASINDIYALTKED 242
KGKSIILMSTHVLDSAB+MCDRFVILH G+RA GTL +L+ FGD +A+LNDIY+ALTKED
Sbjct: 181 KGKSIILMSTHVLDSABRMCDRFVILHKGIVRAQGTADLQRAFGDRSASINDIYALTKED 240

SEQ ID 3454 (GBS353) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 2; MW 30kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 6; MW 55kDa).

GBS353-GST was purified as shown in Figure 216, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1118

A DNA sequence (GBSx1193) was identified in *S. agalactiae* <SEQ ID 3457> which encodes the amino acid sequence <SEQ ID 3458>. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1475 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1119

A DNA sequence (GBSx1194) was identified in *S.agalactiae* <SEQ ID 3459> which encodes the amino acid sequence <SEQ ID 3460>. Analysis of this protein sequence reveals the following:

Possible site: 44

15 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.68	Transmembrane	57 - 73 (50 - 80)
INTEGRAL	Likelihood = -8.49	Transmembrane	122 - 138 (103 - 152)
INTEGRAL	Likelihood = -6.58	Transmembrane	319 - 335 (308 - 337)
INTEGRAL	Likelihood = -4.99	Transmembrane	252 - 268 (249 - 273)
INTEGRAL	Likelihood = -4.19	Transmembrane	104 - 120 (103 - 121)
INTEGRAL	Likelihood = -3.50	Transmembrane	231 - 247 (229 - 248)
INTEGRAL	Likelihood = -1.91	Transmembrane	298 - 314 (298 - 314)
INTEGRAL	Likelihood = -1.44	Transmembrane	28 - 44 (27 - 44)

----- Final Results -----

25 bacterial membrane --- Certainty=0.6074 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12845 GB:Z99109 ABC transporter (membrane protein) [Bacillus subtilis]
 Identities = 101/409 (24%), Positives = 187/409 (45%), Gaps = 76/409 (18%)

35 Query: 1 MKKLFNNKRRSLFLONSKYLYVFNDFVLVIMFLSOLFLLYQYSOLKIDFPKTHWPIVI 60
 M ++ R ++ ++ Y++NDH V+VL+FY8+ ++D P H+P +
 Sbjct: 4 MLDINQSLRQEHIKETRTYKMYMLNDHLVTLVIFFLAGAASWYSKWIRIP-AHPPSPFW 62

40 Query: 61 VSIITMLAMGGIASYLEPADKQFLIKKEAIKETINSKRTYI----- 106
 +++ ++L ++ +L+ AD FL E ++ + A ++Y+
 Sbjct: 63 MAVLPSLVLTSSYKTLKKEADLVFLFLPLKMEPYLKQAFVYSVQSLFLIALSIVAM 122

Query: 107 --FWLVIQTLFLVLISFILIKGL----- 128
 ++ V LV ++ ++L L
 Sbjct: 123 PLYFAVTIPGASLYSVAVFVQLLLKAWNVMBWRTTFQNDRSMKRMVDVIRPAANTLVL 182

45 Query: 129 -----SVFMITLLIFGLGIKILVITYKVKVFYNNQNLINWDAAINHQSRQSLKFPFL 183
 SV+M LL++ + ++L++ K + W++ I E RQ + ++L
 Sbjct: 183 YFVQSVQMYALLVYVIMAVLYLYMSSAAK---RKTFFWESHIESLRRKQFYRIANI 238

50 Query: 184 PTNVKIGISTVKKRSLFDGLILKISKTPSLRWTFNLVFRALRSSDYGLGLTIRLVINILS 243
 PT+V + KRR++LD +L+L+ + + +F RALRSSDYLG+ +RL + L
 Sbjct: 239 FTDVPHLRQAKKRAYLDPLRLRVFPFQRKTAYMPTRAFTRSSDYGLGLVRLITVIFALI 298

55 Query: 244 VIFVNETYIALALAFVFN-YLLLPQLLALGHHPDYQYMNQLYPVRLISAKASQLKQFLRL 302
 +++V+ +L A+ VF ++ QLL L HPD+ + +LYPV+ K ++LK + +L
 Sbjct: 299 IMYVSASPLIAAVLTVFAITFIGIQLLPLFGHPDHALQLYVQV---KETLKRYSPLL 355

Query: 303 SYAVTVIDSI-----LIRELKPVILLIVIMLIVFRYYIPIYKIK 341
 A+++ + L L +I VL+ +V Y+ ++KK

-1250-

Sbjct: 356 KTALSIQALLMSVASAYAAAGLTGFLYALIGSAVLIFVVLPAYMTRIKK 404

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3461> which encodes the amino acid sequence <SEQ ID 3462>. Analysis of this protein sequence reveals the following:

5 Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -14.91 Transmembrane 126 - 142 (119 - 151)
 INTEGRAL Likelihood = -9.77 Transmembrane 320 - 336 (311 - 339)
 INTEGRAL Likelihood = -6.37 Transmembrane 59 - 75 (53 - 79)
 10 INTEGRAL Likelihood = -4.94 Transmembrane 28 - 44 (22 - 47)
 INTEGRAL Likelihood = -4.73 Transmembrane 250 - 266 (249 - 273)
 INTEGRAL Likelihood = -4.04 Transmembrane 231 - 247 (229 - 248)
 INTEGRAL Likelihood = -3.19 Transmembrane 298 - 314 (295 - 315)
 15 INTEGRAL Likelihood = -2.28 Transmembrane 103 - 119 (103 - 119)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6965 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB12845 GB: Z99109 ABC transporter (membrane protein) [Bacillus subtilis]
 Identities = 96/403 (23%), Positives = 173/403 (42%), Gaps = 78/403 (19%)
 25 Query: 1 MKALFLKRRDPQKQNKYLRYVINDHFVILVIMPLGFMVQYGGQLIN---HFPT--- 52
 M ++ R Q+ K+ Y++Y+LNDE V+VL+ F L A Y+ + HPP+
 Sbjct: 4 MLDIWQSRLOQSHIKETRTYMKYLNDEHLVILVIFLPLGASASWYSKMIRDPAPHPFPMVM 63
 Query: 53 -----NILEPQVCLGILIPLLSM----- 71
 L + L L+PL M
 30 Sbjct: 64 AVLFSVLVTSSYVRLTKEADWFLPLLEAKMEPYLQKAFVYSYVSQFLPLIASIVAMP 123
 Query: 72 -----GSTATYLEEADQHFLLPKKEEVI SYI-----KQAERLSFLMAGTLQTAVLL 117
 S+ +Y Q LL +V+ + + +R+ ++ T VL
 35 Sbjct: 124 LYFAVTPGASLVSYYAVFVQLLLKAWQVNEKRTFQNDRSKMKMDVIRFAANTLVLY 183
 Query: 118 FLYPIFRRLGSLFIFILVLILLALKRVVLRSKTRYFLRGNELDNKAVAFESNRKQSI 177
 F++ S+++ +LV +++A+ + +S + W + E RKQ
 Sbjct: 184 FVFQ-----SVMYALLVYVIMAVLVLYMSSAAKR---KTFKWSHIESLELRKQRF 232
 40 Query: 178 LKFYSLEFTTVKGI STKVKERTYINPLLLKVKQTPSNLWLSYARAFRSSDYGLGLFIRM 237
 + +LFT V + + K R YL+ LL+LV + ++ RFLRSSYVLG+ +RL
 Sbjct: 233 YRIANLPTDVHLRKQAKRAYDLFLRLVFPFQRKTFAYMPTRAFRLRSSDYGLGLVRLT 292
 45 Query: 238 LLSSLSVFFIHNILVLSGLALIPN-VLVVFPQLLSLYHYHDYHMTSLYPENSRKKQIML 296
 ++ +L + ++ L ++ +F ++ QLL L+ H+d+ + LYP +K K+
 Sbjct: 293 IVFALIDIMYVASPLIAAVLTVFAIFITIGQLPLFGHFDFHALQELYPQKRETKLKYSF 352
 Query: 297 SFLR-GLSFMLIVNMLCCSSAPKA--LILIVGMVFIACIYLP 336
 S L+ LS L++++ +A L ++G + + LP
 50 Sbjct: 353 SLKATLSIQALLMSVASAYAAAGLTGFLYALIGSAVLIFVVL 395

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/344 (49%), Positives = 237/344 (68%)
 55 Query: 1 MKKLFNKRRLSFLTPQNSKYLRYVINDHFVILVIMPLSGFLLYQYSQKLQDFPKTHWPIIVI 60
 MK LF KRR F Q +KYLRYV NDHFVILVIMPL GE + QY QLL PP H PI V
 Sbjct: 1 MKALFLKRRDPQKQNKYLRYVINDHFVILVIMPLGFMVQYGGQLINHFPTNNHLPQVC 60
 60 Query: 61 VSIIIIIMLIMAGSIASYLEPADKQFLIKKEISAKRTTIFPWLAVIQTLFVLIS 120
 + I+I +LL+MG IA+YLE AD+ FLL KEE + I A++ +++ W +QT L+ +
 Sbjct: 61 LGILIPILLSMWSIATYLEEADQHFLLPKKEEVI SYIQAEKRLSFLMAGTLQTAVLLFLY 120
 Query: 121 PILIKGLSGVPMITLLIPGLGIITKVLVITYKVKVFVNNQNLNWDAAINNEQERKQSLKF 180

-1251-

```

      PT +LGLS+F+ +L+ L +K +V++ K + F      L+W A+ E RQQSILKF
Sbjct: 121 PTFPRRLGSLFTPIILVLLILLAKRVVLSRKTRYFLRGNRLDWAQAVAFESNRKQSLKF 180

5  Query: 181 FSLFTNVKGISTSVKRSFLDGILKLISKIPERLWNLVFRAPLASSDYVLGLTIRLVTLN 240
      +SLFT VKGIST VK R++L+ +LKL+ +TPS LW +L+ RAFLRSSDYVLGL+RL+ L+
Sbjct: 181 YSLFTTVKGIISTKVKERTYLNPLKLKVKQTPENLMLSLHARAFLRSSDYVLGLFLRLMLLS 240

      Query: 241 ILSVIFVNETYLAALAFVFNLLLPQILLALGHHPDYQYMNQLYPVRLNARASQLKGFRLR 300
      LSV F++ YL++LA +FNLL+PQLL+L +H+DY IM LYP +K + FLR
10  Sbjct: 241 SLGVFFIHNLVLSVSLALIFNLYLVFQLGLGLYHYHDHYHMSGLYFENSRKKNLWLSFLR 300

      Query: 301 VLSYAVTVIDSILIRELPVILLIVIMIVTEYYPIPKKKMID 344
      LG+ + +++ + ++LIV M+ + Y+PYK+KK+ID
15  Sbjct: 301 GLSLMLIVINMLCCSSAPKALLIVGMVFIACIYLPYKLLKIID 344

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1120

A DNA sequence (GBSx1195) was identified in *S.agalactiae* <SEQ ID 3463> which encodes the amino acid sequence <SEQ ID 3464>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25  bacterial cytoplasm --- Certainty=0.2821(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:AAC00284 GB:AF008220 YtmP [Bacillus subtilis]
      Identities = 69/214 (32%), Positives = 121/214 (56%), Gaps = 1/214 (0%)

      Query: 12 FLRGKSGKAYIGTYPNGRVSFVKYNTTPIPLPALAKEBQIAPQLIWARFTSNGDMMSAQEWL 71
      P G +G AY + NG+++F+K N++P L L+ E I P+L+W +R NGD++AAQ W+
35  Sbjct: 20 PAGGATGDAYYAQH-NGQQLFLKNGSSPFLAVLSABGIVPKLVYTKRMENGSDVITAQHWM 78

      Query: 72 DGRTLTKEDMGSKQIHHILLHLHKSRLPLVNQLLGLGYKIENFYDLLMDWEKQCTPIQIREN 131
      GR L +DM ++ +L +HH S+ L++ L +LG + NP LL ++ + +
40  Sbjct: 79 TGRLEKPKDMSGRPVABELRKIHITSKALLDMLKRLGKEPLNPGALLSLQKQAVFAVQSS 138

      Query: 132 TYLQSGIVTELRKSLPEFRTTEVATIVBGDIKHSNVIITSGLIYLVMDSVRLTDRMYDVA 191
      +Q + L+ L E + H D+ H+NH+++ +YL+DMD + D D+
40  Sbjct: 139 FLIQSGIKYLEEBHLHEVHPGEKVVCHCQVNHNNWLLSEDNQLYLLIDWGAMTADPAMGL 198

45  Query: 192 YILSHYIPQKHWKDWLSYGYKINEKWSKLIWY 225
      +L HY+ + W+ WLS YG + E + ++ WY
Sbjct: 199 FLNHYIVEKPAWESWLSMYGIELTSLRLRMAMY 232

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3465> which encodes the amino acid sequence <SEQ ID 3466>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55  bacterial cytoplasm --- Certainty=0.2686(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-1252-

Identities = 214/262 (81%), Positives = 242/262 (91%)

Query: 1 MTLSNQLTLTPLRGKSGKAYIGTYPNIGSERVFKYNTTPILPALAKHQIAPQLLWARRTS 60
 +T + QRLTTLPLRGKSGKAY GTYPNIG VP+K NTPILPALAKHQIAPQLLWARRTS
 5 Sbjct: 1 VTTTEQLTLTPLRGKSGKAYGTYPNIGSCVFIKNTTPILPALAKHQIAPQLLWARRTS 60

Query: 61 NGDMMSAQEMLDGRRTLT/KEDMMSKQIHIHLLRLHKSRLVNLQLQLQYKIENPYDLNDV 120
 NGDMMSAQEML+GRRTLT/KEDM SKQIHIHLLRLHKS+ LVNQLQLQL QKIENPYDLN+D+
 10 Sbjct: 61 NGDMMSAQEMLNGRTLT/KEDMMSKQIHIHLLRLHKSRLVNLQLQLQYKIENPYDLNDV 120

Query: 121 EKQTPIQIRENTYLSIVTELKRSLEPFRTEVATVHGDIKHSNWITTSGLIYLVNDWS 180
 E+ P+QI++N+YLQ+IV ELKRSLEP++EVATVHGDIKHSNWITTSGLI+LVNDWS
 10 Sbjct: 121 EQNAPLIQIQNSYLQAIIVKELKRSLEPFRTEVATVHGDIKHSNWITTSGLIYLVNDWS 180

Query: 181 VRLTRMYDVAYILSHYIPQKHWKDWLSTYGYKDNKHWKSWKIIVYQGPSYLSQIICKPDK 240
 VRLTRMYDVAY+LSHYIP+ W +WLSYGYK+N+KV KIIVYQGPS+L+QI+KCPDK
 15 Sbjct: 181 VRLTRMYDVAYLLSHYIPSRWSEWLSYGYKNDKHWKSWKIIVYQGPSHLTQILCKPDK 240

Query: 241 RDMHVNQEIYELKRFREIFRK 262
 RDMHVNQEIY ELKRFREIFRK +K
 20 Sbjct: 241 RDMHVNQEIYELKRFREIFRK 262

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1121

A DNA sequence (GBSx1196) was identified in *S. agalactiae* <SEQ ID 3467> which encodes the amino acid sequence <SEQ ID 3468>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4529 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 35

The protein has homology with the following sequences in the GENPEPT database.

>GP1AAC00285 GB:AF008220 YtmQ [Bacillus subtilis]
 Identities = 126/211 (59%), Positives = 161/211 (75%)
 40 Query: 1 MRVKKRGAEEHLENPHYVIGNPEAKGRWHRIFGNNRPIHIEVGSKGKAPITGMAEQI 60
 MR+R + A++ L N ISNP + KG+W+ +FGN+NPPIHIEVGS+GKG PI+GWA+QN
 Sbjct: 1 MRMRHKPADDPLAENADIAISNPADYKSKMNVFVGHMNPPIHIEVGTGKGQFISGAKQN 60

Query: 61 PDINYIGIDILQVLSVALDKVLDSGARNIKLLVDGSSLNYPDTGEVDLMYLNPSDPW 120
 PDINYIGI++ SV+ A+ KV DS A++KLL +D +L++ F+ GEV +YLNPSDPW
 45 Sbjct: 61 PDINYIGIELFQSVITVAVQKVKDSQAQNVKLLNIDADTLTDFPGEVGRVRYLVNPSDPW 120

Query: 121 PKKHREKRLTYKTFPLDTYKDLPEQGEIHFKIDNRGLFEYSLSPSQYQNTLKVQWLDL 180
 PKK+HREKRLTY FL Y++++ +G IHFKIDNRGLFEYSLSPSQYQNTLKVQWLDL
 50 Sbjct: 121 PKKHREKRLTYSHFLKCYEVMKGGCSHFKIDNRGLFEYSLSKPSSEYGLLLTYVSLDL 180

Query: 181 HASDYQONIMTSEYERKPSNKGQVIYKVEARF 211
 H S+ + NIMTSEY KFS QQ IYR E +
 55 Sbjct: 181 HSNLEGNIMTSEYERKPSALQPIYRAEVEW 211

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3469> which encodes the amino acid sequence <SEQ ID 3470>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 60

-1253-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3303 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 179/211 (84%), Positives = 193/211 (90%)

Query: 1 MRVRKRKGASEHLRNPHYVIGNFEAKGRWHIFGRNNPILHIVGSSGKGAFITGABQN 60
 MRVRKRKGASEHLRNPHYVIGNFEAKGRWHIFGRNNPILHIVGSSGKGAFITGMA+N
 Sbjct: 1 MRVRKRKGASEHLRNPHYVIGNFEAKGRWHIFGRNNPILHIVGSSGKGAFITGMA+N 60

Query: 61 PDINYIGIDIQLSVLSYALDKVLDSGAKNKKLLVDGSSLSNYFDTGEVDLMVLFNSDPW 120
 PDINYIGIDIQLSVLSYALDKVL 8 N+KLL VDGSSLSNYF+ GEVD+MYLNFSDPW
 Sbjct: 61 PDINYIGIDIQLSVLSYALDKVLASEVFNKLLRVGSSLSNYFDTGEVDMMYLFNSDPW 120

Query: 121 PKKHKRRLTYKTFLOTYKDLPEQGEIHFPTDNRGLFEYSLASFQSGWMLKQWLGL 180
 PK KHKRRLTYK FLOTYK ILPE GEIHFPTDNRGLFEYSLASFQSGWML+Q+WLGL
 Sbjct: 121 PKKHKRRLTYKTFLOTYKDLPEQGEIHFPTDNRGLFEYSLASFQSGWMLKQWLGL 180

Query: 181 HASDYQNMITEYERKFSNKGQVIYRVEAF 211
 HAS+Y+ N+MTEYE KFSNKGQVIYRVEA F
 Sbjct: 181 HASNYSGNVMTYEERKFSNKGQVIYRVEANF 211

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1122

A DNA sequence (GBSx1197) was identified in *S. agalactiae* <SEQ ID 3471> which encodes the amino acid sequence <SEQ ID 3472>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1311 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA006136 GB:AP001515 unconserved protein [Bacillus halodurans]
 Identities = 61/124 (49%), Positives = 81/124 (65%), Gaps = 2/124 (1%)

Query: 2 GGEYVLSILIDKPGGIIIVETIAQLTDVVSFLDITQDPDFPEQYMLRVSSEGLERPLKTA 61
 G D+ L+ ID G+ ED ++++++ S LD + DP + Y LEVSSPG ERPLK
 Sbjct: 33 GKDNLRFVFDSETGVLEDCGKVSERLSEKID--ETDPIEQAYPLEVSSPGARPLKRE 90

Query: 62 EALSNAWGSYINVSILKSIDKVIPEGDLLSFDDGTLTIDYMDKTRHKTVIDIPYQTVAKA 121
 + L + +G + +V+L+ ID K E G+ L F DGETLTI+ KTR KTV IPY VA A
 Sbjct: 91 KDLLRSIGKQVHVTLVEPIDGKALGEGELTEFDGSLTITRIKKIKTRKTVIPIYAKVASA 150

Query: 122 RLAV 125
 RLAV
 Sbjct: 151 RLAV 154

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3473> which encodes the amino acid sequence <SEQ ID 3474>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1254-

bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 101/127 (79%), Positives = 117/127 (91%)

Query: 1 MGQDYVLSTILDKPGGITVEDTAQLTDDVVSPLLIITQDPFPFQYMLVSSGGLERPLKT 60
 MG DY+LSIL+DK GGITVEDT+ LT++SPLIITI PDPPF QYMLVSSGGLERPLKT
 10 Sbjct: 52 MGS DYILSILVDKGGITVEDTSDLTNIISPLIITIDPDPFPFQYMLVSSGGLERPLKT 111
 Query: 61 AEALSNVAGSYINVSLEYKSIDKVKIFSGDILLSFGGETLTIDYMKTRHKHTVDIPYQIVAK 120
 A++L AVQSYINVSLEY++IDRVK+F+GDLL+FDGETLTIDY+DKTRHK V+IPYQ VAK
 15 Sbjct: 112 ADSLKAAGVSYINVSLEYQAIIDRVKVFQGDLLAFDGETLTIDYDKTRHKVNIPIYQAVAK 171
 Query: 121 ARLAVKL 127
 R+AVKL
 Sbjct: 172 VRMAVKL 178

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1123

A DNA sequence (GBSx1198) was identified in *S. agalactiae* <SEQ ID 3475> which encodes the amino acid sequence <SEQ ID 3476>. This protein is predicted to be a utilization substance protein a homolog (nusA). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.5069(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 9565> which encodes amino acid sequence <SEQ ID 9566> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB13533 GB:299112 nusA [Bacillus subtilis]
 Identities = 164/370 (44%), Positives = 251/370 (67%), Gaps = 15/370 (4%)
 40 Query: 4 MSKEMLEAFRILEEKKINKEDIIDAVTESLKGANKRRYQGSBCVIEFNKKADPTVYT 63
 MS E+L+A ILE+EK I+KE II+A+ +L SAYKR + Q+++ ++ N + V+
 Sbjct: 1 MSSELIDALITILEKEQISKRILITATEALISAYKRNFNQANVRVDLNRETSGIRVFA 60
 Query: 64 VREVVEVDFDSRLSLESLKDALISSAYELGDKIRFEESVTSFGRVAQSAKQTIMEKVR 123
 ++VVDEV+D RLKTS+++A I Y +GD + E + +PGR+AAQ+AKQ + +++
 45 Sbjct: 61 RKPVDVEVDYQRLSLEISBEAQIHPEYVMGDVVEVTEVTKDFGRILAAQZAKQVTVORVE 120
 Query: 124 QMRVETVNEYKQHEGEIMTQTVRRFDQRFYVNLGSLAQSHQDQIPGESFKSHMDIV 183
 R V ++E+ E +IMTG V+R D +FIYV+LG +EA L+ Q+P ES+K HD I V
 50 Sbjct: 121 AERGVIYSEFDREEDIMTGIVQRLDNKFIYVSLGKIEALLPVNQMPNESYKPHDIRKV 180
 Query: 184 YVYKVENNPKGVNVFVSRSHPEPIKRIMERKIPREVDTGVEIMSRSREAGDRTKAVVSHI 243
 Y+ KVE KG ++VSR+HD +KR+ E E+PE++DGTVE+ SV+REAGDR+K++VR+
 Sbjct: 181 YITKVEKTTKGPQIYVSRTHGLAKRLEIEVPEIYDGTVELKSVAREAGDRSGISVFTD 240
 55 Query: 244 NSNVDAIGTIVGROGSNIKVISNHPKRVDAKTGLEIPVENIDVIQWVEDPAEPIYNA 303
 + +VD +G+ VG G ++ +++ E ID++ W DP EF+ NA
 Sbjct: 241 DPVDVPGVSGVGPKGQGVQVAIVNLK-----GSKIDIVNWSDFVEFVANA 286

-1255-

Query: 304 IAPASVDMVLFDDDDTKRATVVVPSKLSLAIGRRGQNVRLAAHLTGTYRIDIKSASEYEK 363
 ++P++V V+ ++E+ K TV+VD +LGLAIG+RGQN RLAA LTG++IDIKS ++ +
 Sbjct: 287 LSPSKVLVDVIVNKEE-KATTIVTFDYLSLAIGKRGQNVRLAAHLTGKWDIKSETDARE 345

Query: 364 MEAQELQTEEE 373
 + + EE
 Sbjct: 346 LGIYPRELEE 355

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3477> which encodes the amino acid sequence <SEQ ID 3478>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2074 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 348/380 (91%), Positives = 361/380 (94%), Gaps = 2/380 (0%)

Query: 4 MSKEMLEAFRILEEKHINKEDIIDAVTESLKSAYKRRYQGSSECVIEFNEKKADFTVYT 63
 MSKEMLEAFRILEEKHII+K DIIDAVTESLKSAYKRRYQGSSECVIEFNEK ADF V+T
 Sbjct: 12 MSKEMLEAFRILEEKHIDKADIIDAVTESLKSAYKRRYQGSSECVIEFNEKTDPOVPT 71

Query: 64 VRKVVDVDFSRLEISLKDALAISAYELGDKIRFEESVTEFGRVAAQSAKQTMEIKMR 123
 VRKV+EVFDSRLLEISLKDALAISAYELGDKIRFEESV EFGVAAQSAKQTMEIKMR
 Sbjct: 72 VRKVVEEVDFSRLEISLKDALAISAYELGDKIRFEESVNEFGRVAAQSAKQTMEIKMR 131

Query: 124 QMRVTFNEYKHEGEIMTGTVERFDQRFTYVNLGSLEAQLSHQDQIPGESFKSHMDIV 183
 QMRV FNEYK+HEGEIMTGTVERFDQRFTYVNLGSLEAQLSHQDQIPGE+FKSHD IDV
 Sbjct: 132 QMRVTFNEYKHEGEIMTGTVERFDQRFTYVNLGSLEAQLSHQDQIPGETFKSHDRIDV 191

Query: 184 YVYKVENNPKGVNVFVRSRSHPEFIKRIMEREIPEVFDGTVEIMSVSREAGDRTKVAVRSH 243
 YVYKVENNPKGVNVFVRSRSHPEFIKRIMB+EIPEVFDGTVEIMSVSREAGDRTKVAVRSH
 Sbjct: 192 YVYKVENNPKGVNVFVRSRSHPEFIKRIMBQEIPEVFDGTVEIMSVSREAGDRTKVAVRSH 251

Query: 244 NSNVDAIGTIVGRGGSNIKKVISNHPKRVDAKTGLEIPVEENIDVQWVSDPAEFITNA 303
 N NVDAIGTIVGRGGSNIKKVIS FHPKRVDAKTGLEIPVEENIDVQWV+DPAEFITNA
 Sbjct: 252 NPNVDAIGTIVGRGGSNIKKVISKHPKRVDAKTGLEIPVEENIDVQWVDPAEFITNA 311

Query: 304 IAPASVDMVLFDDDDTKRATVVVPSKLSLAIGRRGQNVRLAAHLTGTYRIDIKSASEYEK 363
 IAPASVDMVLFDDDD KRAITVVVPSKLSLAIGRRGQNVRLAAHLTGTYRIDIKSASEY++
 Sbjct: 312 IAPASVDMVLFDDDDTKRATVVVPSKLSLAIGRRGQNVRLAAHLTGTYRIDIKSASEYDR 371

Query: 364 MEAQELQTEEEVASEVISD 383
 +EA+ + A E V+ D
 Sbjct: 372 LEAR--KEAATAVEEPVDD 389

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1124

A DNA sequence (GBSx1199) was identified in *S.agalactiae* <SEQ ID 3479> which encodes the amino acid sequence <SEQ ID 3480>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2012 (Affirmative) < succ>

-1256-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB13534 GB:Z99112 alternate gene name: ymoB-similar to
hypothetical proteins [Bacillus subtilis]
Identities = 46/92 (50%), Positives = 67/92 (72%), Gaps = 1/92 (1%)
Query: 1 M A R T K K I P L R K S V V S G E V I D K R D L L R I V K N K E G Q V F I D P T G K N G R G A Y I K L N D E A I L A 60
M K K K I P L R K V V + G E + K + L + R + V + + K E G + + D P T G K + N G R G A Y + L D + + A
10 Sbjct: 1 M N K H K K I P L R K C V V T G E M K P K E L I R V V R S K E G E I S V D P T G K N G R G A Y I L D K R C I L A A 60
Query: 61 K K K R V F D R S F S M E V S D E F Y D E L L A Y V D H K V K R 92
K K K F + + D + D E L L + K V K +
15 Sbjct: 61 K K K N T L G N O F O S Q I D D Q I F D E L L E L A E - K V K K 91

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3481> which encodes the amino acid sequence <SEQ ID 3482>. Analysis of this protein sequence reveals the following:

20 Possible site: 61
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.1006 (Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/98 (78%), Positives = 92/98 (93%)
30 Query: 1 M A R T K K I P L R K S V V S G E V I D K R D L L R I V K N K E G Q V F I D P T G K N G R G A Y I K L N D E A I L A 60
M + K K K I P L R K S + V S G R + I K R D L L R I V K + G Q V F I D P T G K N G R G A Y I K L N E A + + A
Sbjct: 2 M S R V K K I P L R K S L V S G R I I A K R D L L R I V K T K D G Q V F I D P T G K N G R G A Y I K L N Q E A L M A 61
Query: 61 K K K R V F D R S F S M E V S D E F Y D E L L A Y V D H K V K R R E L G L R 98
K K G + V F + R S F S M + + F Y D + L + A Y V D H K + K R R E L G L +
35 Sbjct: 62 K K K Q V F N R S F S M D I P R S F Y D D L I A Y V D H K I K R R E L G L D 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1125

A DNA sequence (GBSx1200) was identified in *S.agalactiae* <SEQ ID 3483> which encodes the amino acid sequence <SEQ ID 3484>. This protein is predicted to be probable ribosomal protein in infb 5'region. Analysis of this protein sequence reveals the following:

45 Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAR06133 GB:AP001515 unknown conserved protein [Bacillus halodurans]
Identities = 46/95 (48%), Positives = 65/95 (68%), Gaps = 1/95 (1%)
Query: 6 K V I A N L I G L A Q R A G R L I T G E E L V I K A I C N O O V S L I F L A N D A G P N L T K V T D K S N Y K Y E V S 65
K L + L + G L A R A + L + T G E E V + K A + Q N Q V + L + L + + D A G + K K + D K Y + V

-1257-

Sbjct: 5 KWLSELLGLAARARQLITGEEQVVKAVQNGQVTLVLSSDAGIHTKKLLDKGSGYQIPVK 64

Query: 66 TVFSTLELSDALGK-PRKVVAVADAGFSKIMRTLM 99

V + L A+GK R V+ V DAGFS+K+ L+

5 Sbjct: 65 VVGNRQMLGRAIGHKHERVVGVDAGFSRKLALIL 99

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3485> which encodes the amino acid sequence <SEQ ID 3486>. Analysis of this protein sequence reveals the following:

Possible site: 45

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1950 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

15 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 75/99 (75%), Positives = 88/99 (88%)

20 Query: 1 MNNSEKVLNLIIGLAGRAGRLITGEELVKAIQNQVSLIFLANDAGPNLTKVTDKSNYY 60

+ N E+ + LIG AGRAG+I+GEELV+KAIQ+QV L+FLANDAGPN+TKVTDKSNYY

Sbjct: 1 LTNLERLSSLIGFAGRAGKVISGEELVKAIQHQQVTLVFLANDAGPNVTKVTDKSNYY 60

25 Query: 61 KTEVSTVFSTLELSDALGKPRKVVAVADAGFSKIMRTLM 99

EVSTV + LELS ALGKPRK V A+ADAGFSKIMRTLM

Sbjct: 61 NVEVSTVLNLEISALIGKPRKVAIADAGFSKIMRTLM 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1126

A DNA sequence (GBSx1201) was identified in *S.agalactiae* <SEQ ID 3487> which encodes the amino acid sequence <SEQ ID 3488>. Analysis of this protein sequence reveals the following:

Possible site: 37

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2873 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10959> which encodes amino acid sequence <SEQ ID 10960> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3489> which encodes the amino acid sequence <SEQ ID 3490>. Analysis of this protein sequence reveals the following:

45 Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2985 (Affirmative) < succ>

50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 735/961 (76%), Positives = 805/961 (83%), Gaps = 42/961 (4%)

55

-1258-

Query: 1 MSKRLHRLAIKEIGKTSKEVVEQASGLLPVKSHASSVEENDATRIVESPS-SSTKAPT 59
+SKRLHRLAIKEIGK+SKVEVE A+ LGL VKSHASSVEE DA +I+ SFS +SK
Sbjct: 1 LSKRLHRLAIKEIGKSSKEVVEHAKYGLDVKSHASSVERADAKCI ISSPSKASKPDVTA 60

5 Query: 60 NSVQTMQGVKTESKIVETKQGLGDDKSTQPAVKHKPQSRNPKAREERAKAKAEKQIN 119
+ + V S TV + G S+ TQ V+KPK SRNPKAREERAK +A +Q N
Sbjct: 61 SQTVKPKVAPQSPVTVVVKETG-SERVEKIQ-VSKPK--SRNPKAREERAKQAARKQAN 116

10 Query: 120 GD-----HKIKNHNDRSDDDR--HQQKRSKGNR-----NDNRQ-G 154
G +R+ N H D+R H+ Q +N R +DM Q G
Sbjct: 117 GSSHSQRERGQVQPMNHQTNQEGDKRITHRSQDQTNDRKIERKASVNSPRHNDHQLVG 176

15 Query: 155 CQNN---RNKNDGYADHKQKPCPTRPQQAQGRIDPKARAALAKAQNAKYSHSREGRF 210
+N N +GR+ K+ + PQ + +IDPKARAALAKAQNAEYSR SE RF
Sbjct: 177 DRNRPFAKENHIGRPTNQKQGRQBPQSKSP-KIDPKARAALAKAQNAEYSRQSETRF 235

20 Query: 211 REBQAKRAQAKQELAKAALAKQAEAKAKELASKPVAKVKEIVNKVAATPSQTADS 270
R +QEAQR A ++ AK AALKQ E +E A K + + + + TAD+
Sbjct: 236 RAQQEAKRLAEARQAEAKAALAKAASENSHREA-ALKSIEEAEITKLKSNISAKSTADN 294

25 Query: 271 RRKQTRSDKSRQFSNEMEDQKQTPNKKNWNQNVNRQNSMWNHKNHKNKKG---T 326
RRKQ R +K+R ++ +++GQK +NKK+WN+QNVQRN+NSNWN NKK KKKG T
Sbjct: 295 RRKQARPEKNREITHSQEGQK--KNKSNWSNQVNRQNSNWNKNKKTCKGKNVNT 352

30 Query: 327 NGAPKPVTERKFHELKPEFEYTGMTVAELAKRIKREPARIKVKLFPMQVMATQNSIDG 386
N APKPVTERKFHELKPEFEYTGMTVAELAKRIKREPARIKVKLFPMQVMATQNSIDG
Sbjct: 353 NTAQKPVTERKFHELKPEFEYTGMTVAELAKRIKREPARIKVKLFPMQVMATQNSIDG 412

35 Query: 387 DTIELLMVDYGLIEAHAKVEDEADIERFFADEYLNPNLTERPFPVTIMGRVHDKITL 446
DTIELLMVDYGLIEA AKVEVD+ADIERFF DE+YLN+P+ ER PVTIMGRVHDKITL
Sbjct: 413 DTIELLMVDYGLIEAHAKVEVDADIERFFEDENYLNPNLTERPFPVTIMGRVHDKITL 472

40 Query: 447 LDTLNRSRVATGEAGGITQIHIGAYQIEEAGKKITFLDTPGHAAPTSMRARGASVDTITL 506
LDTLNRSRVATGEAGGITQIHIGAYQIEEAGKKITFLDTPGHAAPTSMRARGASVDTITL
Sbjct: 473 LDTLNRSRVATGEAGGITQIHIGAYQIEEAGKKITFLDTPGHAAPTSMRARGASVDTITL 532

45 Query: 507 IVAADDGVMPCTVEAINHNSKAGVPIIVAINKIDKPGANPERVIELAEHGVISTAWGGE 566
IVAADDGVMPCT+EAINHNSKAGVPIIVAINKIDKPGANPERVI+ELAE+G+ISTAWGGE
Sbjct: 533 IVAADDGVMPCTIEAINHNSKAGVPIIVAINKIDKPGANPERVIELAEYGIISTAWGGE 592

50 Query: 567 SEFVEISAKFGKNIDELLETVLVLAEMELKADADVRAIGTVIEARLDKGGKAVATLLVQ 626
EFVEISAKF KNI ELLETVLVLA+BELKAD VRAIGTVIEARLDKGGKGA+ATLLVQ
Sbjct: 593 CEFVEISAKFNKIDELLETVLVLAEBELKADPTVRAIGTVIEARLDKGGKGAATLLVQ 652

55 Query: 627 QGTLNVQDPIVVGNTPGRVAMNDLGRVVKVAGSPFVSITGLNEAPMAGDHFAVAYDE 686
QGTL+VQDPIVVGNTPGRVAM NDLGRVK A PSTFVSITGLNE PMAGDHFAVAYDE
Sbjct: 653 QGTLNVQDPIVVGNTPGRVAMNDLGRVVKVAGSPFVSITGLNETPMAGDHFAVAYDE 712

60 Query: 687 KAARAAGEERAKRALLKQRCNTQVRSLEHFDTLKAGEVKSVMNVIKADVQGSVEALAS 746
KAARAAGEER+KRALKQRCNTQVRS+HLFDTLKAGE+K+VMNVIKADVQGSVEALAS
Sbjct: 713 KAARAAGEERSKRALKQRCNTQVRSLEHFDTLKAGEIKTVNVIKADVQGSVEALAS 772

Query: 747 LKIDVGVGVKVVNVHSAVGAINESDVTIARASNVAIIIGPNVRPTQARQQADDDVEIR 806
L+KI+VBGV+VNVHSAVGAINESDVTIARASNVAIIIGPNVRPTQARQQAD DDVEIR
Sbjct: 773 LVKIDVGVGVKVVNVHSAVGAINESDVTIARASNVAIIIGPNVRPTQARQQADDTIDVEIRL 832

Query: 807 HSIIYKVIIEEVEAMKGLDPYQKILGAIIRSTFKVSKVGTIGGFMVINGKVTIRDS 866
HSIIYKVIIEEVEAMKGLDP YQKILGAIIRSTFKVSKVGTIGGFMVINGKVTIRDS
Sbjct: 833 HSIIYKVIIEEVEAMKGLDPVQKILGAIIRSTFKVSKVGTIGGFMVINGKVTIRDS 892

65 Query: 867 VRVIRDSVVIIDGKLASLKHYYKDDVKVEGNAQGGIMISNNDLKEDDTIAYIMSEIRK 927
VRVIRD VVIIDGKLASLKHYYKDDVKVEGNAQGGIMISN+NDLK DDTIAYIMSEIRK
Sbjct: 893 VRVIRDSVVIIDGKLASLKHYYKDDVKVEGNAQGGIMISNDFNDLKDDDTIAYIMSEIRK 953

65 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1127

A DNA sequence (GBSx1202) was identified in *S.agalactiae* <SEQ ID 3491> which encodes the amino acid sequence <SEQ ID 3492>. This protein is predicted to be ribosome binding factor A (rbfA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 56
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.2557(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9567> which encodes amino acid sequence <SEQ ID 9568> was also identified.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3493> which encodes the amino acid sequence <SEQ ID 3494>. Analysis of this protein sequence reveals the following:

```

    Possible site: 60
    >>> Seems to have no N-terminal signal sequence

20  ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.4765(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

25 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 93/117 (79%), Positives = 103/117 (87%)

    Query: 8  LIMANHRIDRVGMSIKREVNSILRLRVNDPRVQDVITITDVOMLGDLMAKVYFITBSTLA 67
              + MANHRIDRVGMSIKREVNSILRLRVNDPRVQDVITITDVOMLGDLMAKVYFITBSTLA 67
30  Sbjct: 1  MAMANHRIDRVGMSIKREVNSILRLRVNDPRVQDVITITDVOMLGDLMAKVYFITBSTLA 60

    Query: 68  SDNQKAQTGLEKATGTIKRELGNLTMYKIPDLQFVKDESIEYGNKIDMLNLDKK 124
              SDNQKAQ GLEKATGTIKRELGNLTMYKIPDL F KD SI YGNKID+LR+LD K
35  Sbjct: 61  SDNQKAQTGLEKATGTIKRELGNLTMYKIPDLVPEKINSIAYGNKIDQLLRLDNRK 117

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1128

40 A DNA sequence (GBSx1203) was identified in *S.agalactiae* <SEQ ID 3495> which encodes the amino acid sequence <SEQ ID 3496>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

```

    Possible site: 28
    >>> Seems to have a cleavable N-term signal seq.

45  ----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA79277 GB:M64783 acetyl-hydrolase [Streptomyces hygroscopicus]
    Identities = 58/220 (26%), Positives = 90/220 (40%), Gaps = 8/220 (3%)

```

-1260-

Query: 98 WNDGKANQKTIPLYLAGGSYLNPTYPHISMLKTLSTSLDAKILLPIYKTPRYTYDYAI 157
 W ++ +T+ YL GGSY H + L + A++ Y + P + A+
 Sbjct: 58 WVRPARQDGRITLLYLAGGSYALGSPQSHRHLSSALGDAAGAVLALHYRPPESPFPRAV 117

5 Query: 158 PRVLNLYRHPIEKN--ANLTMGDSAGGLALGLAHALSHQSGQRAIPQPMIILLSPW 214
 V YR E+ +TL GDSAG GLA+ AL P P + +SPW
 Sbjct: 118 EDAAVAYRMILLGQCPGPRVTLAAGSAGGLAVALQALR---DAGTFLPAAAVCISFW 173

10 Query: 215 LDVTMKHPEIPKYEDTDPILSNAGLARVGELWANGSNNTNHTYVSPKNAAPATKLAPEITLF 274
 D+ + + + +% I R + G+ + H SP + T L P+ +
 Sbjct: 174 ADLACGASHHTIRKAREILLDTADLRMAERYLAGT-DPRPLASPMKJDTGLPFLLIQ 232

Query: 275 TGTREIFFPDIRDYAAQLQAAHNPVNYIAQGMNHVPIY 314
 G+ E+ D R A PV + M HV+ Y
 15 Sbjct: 233 VGSEVLHDDARALEQAALGACTPVTFEWEPMPIVWHWY 272

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3497> which encodes the amino acid sequence <SEQ ID 3498>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/334 (73%), Positives = 280/334 (83%), Gaps = 6/334 (1%)

30 Query: 1 MKPSFKKILLILLSITILSIACCTPHAKASGRSWSKSWPIRQVFWLKRDKSYKVOEDSSPQ 60
 +K +K L+ ++ L+ TP A AS RSWKSWFIEQVFWLKRDKSY QD+ SPQ
 Sbjct: 1 LKHPIRKTLVTLGLLTLCLP-TPVA-ASSRSWKSFWIEQVFWLKRDKSYKQDDPSQ 58

35 Query: 61 KYLNASREQSDKGYLLDMSVNGGLVQERLFDQMVSVDNNGKANQKTIPLYLAGGSYLN 120
 +YL+A REQSDK Y LD N VNG LVQE L+ MVSVDNNGK +QKTI YLAGGSYLN
 Sbjct: 59 RYLDACRQSDKPYQLDNLVNGPLVQENLYGMQVYSVDNNGKPDQKTIYLAGGSYLN 118

Query: 121 PTPYHISMLKTLSTSLDAKILLPIYKTPRYTYDYAI PRVLNLYRHPIEKNANLTMGDS 180
 PT YHI+MLKTLSTSLDAKI+LPIYK PRTY+Y +P+LVNLY+H++ KN N+ LMGS
 40 Sbjct: 119 PTPYHINMLKTLSTSLDAKIVLPYKAPRTYNTYTMPLVNLYQHYHKNQNVFLMGDS 178

Query: 181 AGGLALGLAHALSHQSGQRAIPQPMIILLSPWLDVTMKHPEIPKYEDTDPILSNAGLA 240
 AGGLALGLAHAL + E++PQPK ++LSPWLDVTM HPEIP+YED DPILS+GL
 Sbjct: 179 AGGLALGLAHALN---ESVPQPKQVLSPWLDVTMNSHPEIPPEYEDDPILS+GLK 234

45 Query: 241 RVGEIANGSNNTNHTYVSPKNAAPATKLAPEITLFTGTREIFFPDIRDYAAQLQAAHNPV 300
 RVGE+WA ++NMH VYSPK P T L PITLFTGTREIFFPDIRDYAA+L+AMH +
 Sbjct: 235 RVGELWAYSDNNTNHTYVSPKNGEITLPTLITLFTGTREIFFPDIRDYAAKLNANNT 294

50 Query: 301 YIAQEGMNHVPIPIPEAKTAQYQMIIDINKTP 334
 +I QEGMNHVPIPIPEAKTAQYQ+ID INKTP
 Sbjct: 295 FITQEGMNHVPIPIPEAKTAQYQIIDINKTP 328

A related GBS gene <SEQ ID 8731> and protein <SEQ ID 8732> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 11.88
 GVH: Signal Score (-7.5): -1.33
 Possible site: 28
 60 >>> Seems to have a cleavable N-term signal seq.
 ALON program count: 0 value: 4.03 threshold: 0.0
 PERIPHERAL Likelihood = 4.03 174
 modified ALON score: -1.31

-1261-

*** Reasoning Step: 3

----- Final Results -----

```

5      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

10 28.4/46.2% over 220aa

Streptomyces

hygroscopicus
EGAD|5925| acetyl-hydrolase Insert characterized

15 ORF00486(589 - 1245 of 1602)
 EGAD|5925|5724(57 - 277 of 300) acetyl-hydrolase {Streptomyces hygroscopicus}
 %Match = 6.8
 %Identity = 28.3 %Similarity = 46.1
 Matches = 62 Mismatches = 111 Conservative Sub.s = 39

[illegible][illegible]

927 957 987 1017 1047 1077 1107 1137
RAIPQPKNIILLFSLDVTGMKHPPEIKPYEDTDPILSAMGLARKGEIWAINGSNNHTVVSPPKNAPATKIAPITLFTGTRE
| | : : ||| | : : : : : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
AGTLPAAAVACISFWADLCGASHTRKARELLDTADLRMAERYLAGTD-PRHPLASPAHGDLTGFLPPLLQVGSE

170 180 190 200 210 220 230

40 1167 1197 1227 1245 1275 1305 1335 1365
IFPFDIRDYTAQIQAAHFHVMNYIAQSNNV----FYIYPPEEAKTAYQMIDIINKPT*Y*LSQL*SYVK*TMLTWFI
| | | | | | | | : | : | : | : |
VLHDARALEQAALGAGTPVFTEEMFMFHVWHYHPVLPEGRRAAIEVAGFLRTATGEGGLK
 250 260 270 280 290 300

45 SEQ ID 8732 (GBS149) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 6; MW 37kDa).

The GBS149-His fusion product was purified (Figure 196, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 291), which confirmed that the protein is immunoaccessible on GBS bacteria.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1129

A DNA sequence (GBSx1204) was identified in *S. agalactiae* <SEQ ID 3499> which encodes the amino acid sequence <SEQ ID 3500>. This protein is predicted to be CopY. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3140 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
Identities = 67/137 (48%), Positives = 98/137 (70%)

Query: 2 TISSAEWEIMRVVWQAQNTTSNAILAVLLEKYDNTPTVTKLLRLLDKGYVSRMGKG 61
+IS+AEWE+MRVVWA+Q T+S+EI+A+L Y W+ ST+KTL+ RL +KGVA+ +G+
Sbjct: 3 SISAWEWEIMRVVWQAQNTTSNAILLGRITVYCSASTIKTLITRLSEKGYLTSSQQRK 62
Query: 62 FYSYPLIDEDLAWMSSEVDSVPQKVCQTKHVAIVRHLESIPMTKDRNLNQLSSLEAKGK 121
+YS LI E+ A+ +V VP ++C TKH A++RHL+E PMT D L++ L+KK
Sbjct: 63 YIYSLISESEALEQQVSEVPSRICVTKHQALIRHLVEETPMTLEDIEKLEALLLEKCAN 122
Query: 122 TLERVACNICIPGQCC 138
+ V CNCI GQC C+
Sbjct: 123 AVFEVKNKNCIVGQCSCY 139

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3501> which encodes the amino acid sequence <SEQ ID 3502>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2331 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/135 (40%), Positives = 84/135 (62%)

Query: 3 ISSAEWEIMRVVWQAQNTTSNAILAVLLEKYDNTPTVTKLLRLLDKGYVSRMGKG 62
IS+AEWE+MRVVWA + S++I+ +L +KY W+ ST+KTL+ RL+ K+++ +G+ +
Sbjct: 10 ISAEWEIMRVVWASGDIKSSDIITILRKQYQMSDSTIKTLIGRLVKQPLTSYRQGRAY 69
Query: 63 FYSYPLIDEDLAWMSSEVDSVPQKVCQTKHVAIVRHLESIPMTKDRNLNQLSSLEAKGK 122
Y L+DE L + +V ++CQ +H ++ L +PMT ++ Q LE KK
Sbjct: 70 IYQALLDEFTALQKEALATVLDGICQNRILLRLYHLPMLEETIGAPRLLEKVENA 129
Query: 123 LERVACNICIPGQCC 137
+ V CNCA+PGQC C
Sbjct: 130 VLEVPKNCLPGQCNC 144

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1130

A DNA sequence (GBSx1206) was identified in *S.agalactiae* <SEQ ID 3503> which encodes the amino acid sequence <SEQ ID 3504>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.82	Transmembrane	382 - 398 (370 - 406)
INTEGRAL	Likelihood = -8.01	Transmembrane	356 - 372 (344 - 374)
INTEGRAL	Likelihood = -2.50	Transmembrane	719 - 735 (719 - 738)

-1263-

INTEGRAL Likelihood = -2.28 Transmembrane 202 - 218 (202 - 218)
INTEGRAL Likelihood = -1.59 Transmembrane 693 - 709 (691 - 712)
INTEGRAL Likelihood = -1.33 Transmembrane 167 - 183 (167 - 183)

5 ----- Final Results -----
bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10086 GB:A296446 CopA [Streptococcus mutans]
Identities = 440/740 (59%), Positives = 571/740 (76%), Gaps = 1/740 (0%)

15 Query: 5 KETFLIDGMYTCASCAITIEKAVNKLEVDASAVVNLAIEKMTIVFDITILSPVWIECVSE 64
+E FLIDGMYTCASCA+ +E AV KLD ++SAVNIL TRMT+ +D +S + + V+
Sbjct: 3 EKVFLIDGMYTCASCAINVENAVKLEGIASAVVNLAIEKMTIVFDITILSPVWIECVSE 62

20 Query: 65 SGYEASLFKEETSKSQSERHQLAIEKMWHRFWMSAVATIPLLYISMGPMINLWLPFLMP 124
+GY A ++ T++SQ +R + + + R +++ TIPL YI+M3 M L LP+FL P
Sbjct: 63 AGYGAKVYDPTTABSQNDREKHEKLAGIKKRLWTSIFITPLFYIANGSMVGLPFLFLAP 122

25 Query: 125 DKGPLANYGMQILLTLFWMYGRIFYKQFKALFKRHFNMDSLVAIATAAFIYISLYGLY 184
FL Y M+ LLTL+PV+ FY NGF++LFK HEMDSLV++ATTAFA+YSLG Y
Sbjct: 123 SSAPLTYAMVLLLLT:PVVILWMSFYDNGFKSLFKGHFNMDSLVSTATTAAFLYISLYGY 182

30 Query: 185 ELIQDIIHYAHQLYFESVAVILTLITLQKYFELSGKRTSASIEKLLTSGAKEARVITDG 244
+ G H+AE LY+BSVAVILTLITLQKYFELSGKRTS +I+KL+LSAKEA +IDG
Sbjct: 183 HVLGHTHHHLYFESVAVILTLITLQKYFELSGKRTSADIKKMLHSAKEATLIRG 242

35 Query: 245 EBYMPLDKVKIGSTILVKGEGKIPLDGHWVAGESSIDESMLTGESIPVEKKVQSGKYGA 304
E+ VP++V++I+ TLVKGEGKIP+DG V++G S+IDESMLTGESIP+EK S YV
Sbjct: 243 EIKVPIEQVQIRIQDILVKGEGKIPVDGRVLSGSAIDESMLTGESIPVEKKVQSGKYGA 302

40 Query: 305 SINGQGSILTFPEKEAGGSLISQIINLVEAAQTSKAPIANLADKVSQVFPFVITVAIL 364
SINGQGSILTF EK +LLSQII LVE AQ +KAPTA +ADKVS VFPF +I IAIL+
Sbjct: 303 SINGQGSILTFPEKEAGGSLISQIINLVEAAQTSKAPIANLADKVSQVFPFVITVAIL 362

45 Query: 365 GLWYLLIQSSPAPSLKIMIAVLVLIACPCALGATPTTAINVWASGKAEMNLPFGGSE 424
GL WY ++OQ P FS+ I +AVLVLIACPCALGATPTTAINVWASGKAEMNLPFGGSE 424
Sbjct: 363 GLWYLLIQSSPAPSLKIMIAVLVLIACPCALGATPTTAINVWASGKAEMNLPFGGSE 422

50 Query: 425 KAHHIDTIVFDKTTOTLTKGKPEVVAIKTYGGDKKEFLQGVASVEKLSNHLSQITVNAK 484
AH I+TIVFDKTTGT+TKGPEVV +Y D+ + + A++LS HPLSQIV AK
Sbjct: 423 LAHQINTIVFDKTTGTITQSGKEVWHQFSY-HDRDVLQVTAALSAEHLQSAIVDYAK 481

55 Query: 485 EKEPLREVEAFPMNKLIVGLSATINGTMLVGNMNTKNDVNLDAKADIRIAQEAQ 544
++ L V F ++ G GL + +T+LVGN LM +++++ A+AD +A +QT
Sbjct: 482 KEETHLLAVDDPTSLTGLGLKGCVADESLVGNMNTKNDVNLDAKADIRIAQEAQ 541

60 Query: 545 VVYSENGVLSGLITLTDQKIDSQRTVQLQIRLGNLVLITGDNKAAPDAIAQKIGIT 604
++V+ +G L GLIT+ D+K DS TVK LQ +G + +LTGDN+ +A AIA++GIT
Sbjct: 542 PIFVASDQQLGLITLTIADKVNDSANTVKALQNMGEVAMLTGDNNEFTQAIAKEVGIT 601

65 Query: 605 VVSEVLDPKAVNILELKEGSGQIAMVGDGINDAPALASSDVGISMSSGGTDAIESADIV 664
V+8+V +K IL+L+ +G +AMVGDGINDAPALA++D+DISM SGTDIA+ESADIV
Sbjct: 602 VISQVPSQEKTA+LDLQA3GKKVAMVGDGINDAPALATADIGISMSSGGTDAIESADIV 661

Query: 665 LMKPELTDLKAMPTIQTITIKENLFWAFFYVNLAIIVAMGVILHGGPLIMFMAAGL 724
LMKP + D++KA+ IS+ TIKENLFWAF YNVL++P+AMGVILHGGPLIMFMAAGL 724
Sbjct: 662 LMKPAMLDIICALKISRVTIINIKENLFWAPIYNVLVSPVIMGVILHGGPLIMFMAAGL 721

Query: 725 AMSFSSSVVNLALRLKVK 744
AM+FSFSSSVVNLALRLKVK+K
Sbjct: 722 AMSFSSSVVNLALRLKVK 741

There is also homology to SEQ ID 3506.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1131

A DNA sequence (GBSx1207) was identified in *S.agalactiae* <SEQ ID 3507> which encodes the amino acid sequence <SEQ ID 3508>. This protein is predicted to be cation-transporting ATPase, P-type (pacS). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1934 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AA010087 GB:AF296446 CopZ [Streptococcus mitans]
Identities = 31/67 (46%), Positives = 43/67 (63%)

Query: 1 MEHTYRVSGMKCDGCAKTVSDKLSVIGVDEVNVDLTINQVVVSGKTFKMLLRSLKDTK 60
      M+ TY + G+KC GCA V+ + S + V++V VDL K +V ++G KM LKR+LK T
Sbjct: 1 MEKTYTHDGLKCGCADNVTKRPSLELKVNDVKVDLDKKEVRITGNPSKMSLKRALGKTN 60

Query: 61 YSLEEEI 67
      Y L EI
Sbjct: 61 YELGAEI 67
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3509> which encodes the amino acid sequence <SEQ ID 3510>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2997 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 33/63 (52%), Positives = 48/63 (75%)

Query: 1 MKHTYRVSGMKCDGCAKTVSDKLSVIGVDEVNVDLTINQVVVSGKTFKMLLRSLKDTK 60
      M+ Y+V+GM CDGCA+TV++KLS+V GV V V+L K + V+G+ +L+KR+LKDTK
Sbjct: 1 MEKHYQVTGTMTCDGCAKTVTEKLSAIPGVQSVQVNLKEGKAIVKRGPLTFILKRALKDTK 60

Query: 61 YSL 63
      + L
Sbjct: 61 FEL 63
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1132

A DNA sequence (GBSx1208) was identified in *S.agalactiae* <SEQ ID 3511> which encodes the amino acid sequence <SEQ ID 3512>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
```

-1265-

```

>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -7.59 Transmembrane 67 - 83 ( 65 - 90)
INTEGRAL Likelihood = -3.72 Transmembrane 35 - 51 ( 31 - 51)
INTEGRAL Likelihood = -3.61 Transmembrane 122 - 138 ( 120 - 139)
5 INTEGRAL Likelihood = -1.59 Transmembrane 154 - 170 ( 154 - 171)

----- Final Results -----
bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8733> which encodes amino acid sequence <SEQ ID 8734> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 5
MoG: Discrim Score: 4.09
GVH: Signal Score (-7.5): 3.87
Possible site: 20
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 4 value: -7.59 threshold: 0.0
20 INTEGRAL Likelihood = -7.59 Transmembrane 65 - 81 ( 63 - 88)
INTEGRAL Likelihood = -3.72 Transmembrane 33 - 49 ( 29 - 49)
INTEGRAL Likelihood = -3.61 Transmembrane 120 - 136 ( 118 - 137)
INTEGRAL Likelihood = -1.59 Transmembrane 152 - 168 ( 152 - 169)
PERIPHERAL Likelihood = 0.85 96
25 modified ALOM score: 2.02

*** Reasoning Step: 3

----- Final Results -----
30 bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35 >GP: CAB15351 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
Identities = 107/192 (55%), Positives = 137/192 (70%)

Query: 8 WNILSLVGTAFASSGAIVAEFEFDILGLFILGFVIAFGGGAIRNVLIQLPIETMSQG 67
W +LS++G +AFA SGAIVA+EEE+DILG++ILG VTAFGGGAIRN+LIG+P+ LW QG
40 Sbjct: 3 WELLSVIGIIAFAVSSGAIVAMEEYDILGVYILGIVTAFGGGAIRNLIIGVPSALMEQG 62

Query: 68 IAFYAAAAILFIMIFPNLLSGKGRDAEVSDAIGLAFSVQGLATQSHQPLSAVIA 127
F A +I + +FP LL +SDAIGLAARF++QGALYA + PLSAVIA
45 Sbjct: 63 AYPQLALLSITIVLEFPKLLLSHMNKGHLSDAIGLAFAIQGALYAVKQGHPLSAVIA 122

Query: 128 AVLITGAGGGIIVRDVLGRKPGVLRSEIYAGNSILVGIILYFKIAKTTIDYXLLVNVVTSL 167
AVLTG+GGGI+RD+LAGRKP VL++EIIYA W+ L G+I+ + Y+L V+
Sbjct: 123 AVLITGSGGGIIRDLLAGRKPLVWKARIYAVNAALGGLVGLGWLGNPSGLYLVFPFVWVC 162

50 Query: 188 RNLGYKKQNHLP 199
R+ Y W LP
Sbjct: 183 RVCSPYMPNKKLP 194

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3513> which encodes the amino acid sequence <SEQ ID 3514>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -5.15 Transmembrane 70 - 86 ( 65 - 88)
INTEGRAL Likelihood = -4.09 Transmembrane 33 - 49 ( 29 - 49)
60 INTEGRAL Likelihood = -2.13 Transmembrane 120 - 136 ( 119 - 137)
INTEGRAL Likelihood = -0.43 Transmembrane 173 - 189 ( 172 - 189)

----- Final Results -----

```

-1266-

bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 The protein has homology with the following sequences in the databases:

>GP:BA05428 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 109/195 (55%), Positives = 137/195 (69%)

10 Query: 6 WEIINIIIGTIAFALSGAIVAMEEFDILGIFILGFTAFGGGAIRNLTIGLPIEALWQK 65
 W++LN+IGTIAFALSG IVAMEE+VD++G+ILGFTAFGGGAIRN LIG+P+ ALW Q
 Sbjct: 3 WDVINIVIGTIAFALSGIVAMEEFDLNGVYILGFTAFGGGAIRNLTIGVPSALWQK 62

Query: 66 PEPTCAFFAMVLIMLPKLMARGWVRVAVLTDALGLAAPSVCALHAVRLNQPLSAVIV 125
 PT AF M + P L W++ +L DAIGLAAF++QCAL A ++ PLASAVIV
 15 Sbjct: 63 TLFTIAFIVMTIAFPLKLNINHLKPGILLFDALGLAFAATQCALPATSMDHPLSAVIVA 122

Query: 126 AVLTGAGGGVVRDILAGRKPSVLRSEIYAGNSILAAIVLHFKLADSTIECYALVLLTTL 185
 A LTGAGGG+VRD+LA RKP VL EII GN++LA + + I , L++L+ L
 Sbjct: 123 AALTGAGGGIVRDLARKKPLVLSKETIYIGNAMLAGAIGLINTVSQGIQGLIILVFL 182

20 Query: 186 RMIGNRKKWNLPKIK 200
 RM+ W LP K
 Sbjct: 183 RMLSVHYNWLPHRK 197

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/200 (66%), Positives = 168/200 (83%)

Query: 3 NSIDINWILSLVGTVAFASSGAIVAEIEEFDILGIFILGFTAFGGGAIRNLTIGLPIET 62
 M+ID+W IL+++GT+AF+ SGAIVA+EEFDILG+FILGFTAFGGGAIRN LIGLPIE
 30 Sbjct: 1 NTIDMWEILNIIGTIAFALSGAIVAMEEFDILGIFILGFTAFGGGAIRNLTIGLPIEA 60

Query: 63 LWSQGIAYAAAAIILFIMIFNNLSGKGRDAEVSDAIGLAAFSVQCALYATQSHQPLS 122
 LW Q F A A++ IM+FF L++ A V++D+IGLAAFSVQCAL+A + +QPLS
 35 Sbjct: 61 LWSQKPEPTCAFFAMVLIMLPKLMARGWVRVAVLTDALGLAAPSVCALHAVRLNQPLS 120

Query: 123 AVIVAALVLTGAGGGIVRDLVLRKPSVLRSEIYAGNSILVGIILYFKIAKTTTIDYLVNL 182
 AVIV AVLTGAGGG+VRD+LAKRKP VLRSEIYAGNSIL I+L+FK+A +T + Y LV+
 Sbjct: 121 AVIVTRVLTGAGGGVVRDILAGRKPSVLRSEIYAGNSILAAIVLHFKLADSTIECYALV 180

40 Query: 183 VVTSLRWLGYKKWNLFPVR 202
 ++T+LKM+G +K+M-LP ++
 Sbjct: 181 LLTLRLMIGNRKKWNLFPKIK 200

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 1133

A DNA sequence (GBSx1209) was identified in *S.galactiae* <SEQ ID 3515> which encodes the amino acid sequence <SEQ ID 3516>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2805(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9569> which encodes amino acid sequence <SEQ ID 9570> was also identified.

-1267-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA894816 GB:AJ245582 hypothetical protein [Streptococcus thermophilus]
Identities = 138/238 (57%), Positives = 184/238 (76%)

5 Query: 5 KMIKLIADMDGTLINDEKKIPKENIQAIRATQAGIKIVLCTGRPMGILPYFNLGL 64
+ KLIADMDGTLIN +K+IPKENI AI+EA AGIKIVLCTGRP SGI+P+P +LGL
Sbjct: 3 QNQVKLIADMDGTLINQKIPKENIKAIRATQAGIKIVLCTGRPSGIVPAPFKLGL 62

10 Query: 65 TKEEYIIMNGCSTYSTKDWQLIDSATLTHDELIFLEVVKEFPNVLTLTAENTFYAVG 124
++EE+IIMNGCSTY TK+W L+AS +L+ E+ L+ ++FP V LT T E +Y+Y VG
Sbjct: 63 SEEFYIIMNGCSTYSTKDWQLIDSATLTHDELIFLEVVKEFPNVLTLTAENTFYAVG 122

15 Query: 125 EEPVEIVAYDADLVFTKAKSTLDALRNQEEIVFQPMYMGLDADVAFQRAVERALISKP 184
EVPE+VAYDA VPT+AK+ SL+ - + ++PQMYN + AFQ AV+ + L +
Sbjct: 123 NEVPVELVAYDAGTVFTKAKSRLEEIPFGQVIFQPMYMAESEPLDAPQNAVQRDLQSY 182

20 Query: 185 SGVRSQDYIEIMPQGVTKARGLSLAKGLDINQVMAIGDAPNDIELLDLVPSNVA 242
S VRSQ+YI+E+MPQG TKA GLK L KL ++ +Q+MA+GDA ND+E+L V SVA
Sbjct: 183 STVRSQSYIEFVMPQGVTKARGLSLAKGLDINQVMAIGDAPNDIELLDLVPSQVA 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3517> which encodes the amino acid sequence <SEQ ID 3518>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1468 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An identity of the GAS and GBS proteins is shown below.

Identities = 152/270 (56%), Positives = 193/270 (71%)

35 Query: 6 KMIKLIADMDGTLINDEKKIPKENIQAIRATQAGIKIVLCTGRPMGILPYFNLGL 65
+MI+LIADMDGTLIN +K+IPKENI AI+EA Q+G+KIVLCTGRP SG PYP+LGLT
Sbjct: 19 RMIQLIADMDGTLINQKIPKENIQAIRATQAGIKIVLCTGRPQSGTRPYDQLGLT 78

40 Query: 66 KEEYIIMNGCSTYSTKDWQLIDSATLTHDELIFLEVVKEFPNVLTLTAENTFYAVG 125
++EE+I+MNGCSTYS+ DWQL S L ++ LEE+ + FP++ LILT EN + + R
Sbjct: 79 QEEFLIIMNGCSTYSTSPDWQLRHSKMLKVSDELLEELQSFPDILVLTLENDYVLIRE 138

45 Query: 126 EEPVEIVAYDADLVFTKAKSTLDALRNQEEIVFQPMYMGLDADVAFQRAVERALISKP 185
EVP++V D DLVFT K SL L + ++PQMY+G A + AF+ AV L F
Sbjct: 139 EYPDLVQEGDGLVFTIKVPSLAELDLTPRIFQPMYLGKALDAFRAVRLQSLQSFH 198

50 Query: 186 GVRSDQDYIEIMPQGVTKARGLSLAKGLDINQVMAIGDAPNDIELLDLVPSNVA 245
VRSQD I EI+POGV+KA LK L- LGL +QVMAIGDAPNDIE+L VM N
Sbjct: 199 VVRSDQNIIEILPQGVSKASALKELVEDLSLTADQVMAIGDAPNDIELMLYGLVPMEN 258

50 Query: 246 ASDIKSRCKYITVDNNKAGVAKAYIDYAL 275
AS IK +T+ N+ AGVA+AI +AL
Sbjct: 259 ASAAIKPLADKVTITDNNAGVAQAIRQVAL 288

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1134

A DNA sequence (GBSx1210) was identified in *S.agalactiae* <SEQ ID 3519> which encodes the amino acid sequence <SEQ ID 3520>. Analysis of this protein sequence reveals the following:

Possible site: 18

-1268-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 7 - 23 { 7 - 23 }

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AA26954 GB:J04479 DNA polymerase I [Streptococcus pneumoniae]
 Identities = 655/879 (74%), Positives = 746/879 (84%), Gaps = 4/879 (0%)
- Query: 3 NKKLLLDGSSVAFRAFFALNQIDRFKMSGLHTNAIVGFHMLAHILGRVQPSHLV 62
 +K KLLLDGSSVAFRAFFALY Q+DRFKN +GLHTNAIVGF IML+H+L RV+PSHLV
 15 Sbjct: 2 DKKLLLDGSSVAFRAFFALYQQLDRFKNAGLHTNAIVGFQMLSHLERVEPSHLV 61
- Query: 63 AFDAGKITTFRTENYADYKGRAKTDPDEFBQFPFYIRQQLDVLGIHYEELHYEADDIIGT 122
 AFDAGKITTFRTENYADYKGRAKTDPDEFBQFP+IR+ LD +GI+HYEL YSADDIIGT
 20 Sbjct: 62 AFDAGKITTFRTENYADYKGRAKTDPDEFBQFPFYIRQQLDVLGIHYEELHYEADDIIGT 121
- Query: 123 LAKQAEASNEHFDITVSGGKDLILQITDITVVEISKKGVAEFPFAYLMSEMGITPS 182
 L K AE + FDIT+VSGGKDLILQITDITVVEISKKGVAEFPFAYLMSEMGITPS
 20 Sbjct: 122 LDKLAE--QDGFITVSGGKDLILQITDITVVEISKKGVAEFPFAYLMSEMGITPS 179
- Query: 183 QFIDLKALMGKSDNIPGVTKIGETGLKLLSEYGLSGLIYENIRAMKSKKKNLINDK 242
 QFIDLKALMGKSDNIPGVTK+GKTK+KLL E+GSLGIIYENI+ MK SKKKNLINDK
 25 Sbjct: 180 QFIDLKALMGKSDNIPGVTKIGETGLKLLSEYGLSGLIYENIRAMKSKKKNLINDK 239
- Query: 243 EQAFLSKTLATINIASPTITGLEDILVYSGPDIALGQFYDEMDFQKPTALGSESTQED 302
 EQAFLSKTLATI+ +PI IGLED+VYSGP D++ L +FYDEM FQK K AL ++
 30 Sbjct: 240 EQAFLSKTLATIDTKAPIAGLEDILVYSGP-DVENLKGFPYDEMGFKQLGALWSSADVA 298
- Query: 303 FEVDPTVEVQLKTEFSDNDFFYFEMLDGNHVEDLIGANGNSDITVSNVSLQAL 362
 +DPT V+Q+ +H S+ ++FE+ Q+NYH ++L+G AN D +YAT + LIQ+ +
 35 Sbjct: 299 EELDPTIVDQISQDMLSEESIHPFELPGENYHTDMLGFWAGSODQYATDKLELQDPI 358
- Query: 363 FKALSKP- IKTYDFKRSKVLNRPFHLDPEAFDTRAKYLLSTENLVSTRIATLTH 421
 FK L K ++ YDFK+ KVLV RF +DL PAFD RLAKYLLST EN +TIA LY
 40 Sbjct: 359 PKDFLEKTSKLVYDFKVKVLLQRFGVDLQAPFDIRLAKYLLSTVENRIATIASLVQ 418
- Query: 422 LPLITDADVQKGAIRAIPEKTRFLEHLAKVKVLVDSANIMQOLKANOEELFEME 481
 L D+ YGK K+AIPE+ +FLEHL K+ VLV+AE +++++ N Q ELL+NEQ
 40 Sbjct: 419 TYIVDETFYQKGVKGAIPEREKPLEHLACKLAVIVETPELLEKLENGQLLELYNE 478
- Query: 482 PLANVLAKMEIRGIKVKKNTINEMAIENKQVIETLTQRIEYELAGQFNINSPKQIKLLF 541
 PLA VLAKMEI GI VKK TL EM EN+ VIE LTQRIEYELAG+EFN+NSPKQLG LLF
 45 Sbjct: 479 PLAFVLAKMEIAGIVKKEITLLEMQARENLVIEKLTQRIEYELAGQFNINSPKQLGVLLF 538
- Query: 542 ETGLGPVETTKTKTGYSTAVDVLERLAPISLVTKILEYRQITKLGSTYIIGIQVILE 601
 E LGLP+E TKTKTGYSTAVDVLERLAPI+P+V KIL+YRQI K+GSTY+IGIQD+TL
 50 Sbjct: 539 EKILGLPLEYTKTKTGYSTAVDVLERLAPIPIVKKILDYRQIAKIGSTYIIGIQDMLA 598
- Query: 602 DCKIHTRYVQLTQTRGLSSDPNLQNPVRLPGRGLIRKAFVPSNDNAVLSSDVSQIE 661
 DCKIHTRYVQLTQTRGLSS DPMLQNP RLEGRGLIRKAFVP +++VLLSDVSQIE
 55 Sbjct: 599 DCKIHTRYVQLTQTRGLSSDPNLQNPRLPGRGLIRKAFVPSNDNAVLSSDVSQIE 658
- Query: 662 LRVLAHISKDEHLIAAFKEGADIHTSTAMKRVFGIEKPENVTPNDRNAKAVNFGVYGIS 721
 LRVLAH+SKDEHLI AF+EGADIHTSTAMKRVFGI+P+NV+ NDRNAKAVNFGVYGIS
 60 Sbjct: 659 LRVLAHISKDEHLIKAFQEGADIHTSTAMKRVFGIERPENVTPNDRNAKAVNFGVYGIS 718
- Query: 722 DPLGSLNGLIPKILAKQYIDTYFRYPGIQNYMDEVREARDKGYVETLPHRRSRPLDIN 781
 DPLGSL+NLGI RK AK AYIDTYFR+PGIQNYM+ VUREA+DKGYVETLPHRRSRPLDIN
 60 Sbjct: 719 DPLGSLNGLISRKAKAYIDTYFRYPGIQNYMDEVREARDKGYVETLPHRRSRPLDIN 778
- Query: 782 SRNFNIRQFAETAINSPIQGSADILKIAMINORVLIDGGYKSKMLQVHDEIVLEVP 841
 SRNFNIR FAE TAINSPIQGSADILKIAMI LD+ L GGY+KMLQVHDEIVLEVP
 65 Sbjct: 779 SRNFNIRQFAETAINSPIQGSADILKIAMIQDKALVAGGYQTKMLQVHDEIVLEVP 838

-1269-

Query: 842 NEEIGAIRESLVTKTMSAISLSVPLIADENAGETWYEAK 880

E+ +++LV +TME AI LSVPLIADEN G TWYEAK

Sbjct: 839 KSELVEMKKLVKQTMEEAIGLSVPLIADENEGATWYEAK 877

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3521> which encodes the amino acid sequence <SEQ ID 3522>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

10 INTEGRAL Likelihood = -0.43 Transmembrane 7 - 23 (7 - 23)

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 665/881 (75%), Positives = 761/881 (85%), Gaps = 2/881 (0%)

20 Query: 1 MTHNKKLLLDIGSSVAFRAFPALYNQIDRFKNSGLHTNAYGPHMLNHLGRVOPSHI 60

M NKNKLLLDIGSSVAFRAFPALYNQIDRFKN+SGHATNAYGPHML+H++ RVOP+H+

Sbjct: 1 MENKKNLLLDIGSSVAFRAFPALYNQIDRFKNHSGHATNAYGPHMLDHMKKRVQPHV 60

25 Query: 61 LVAFDAGKTFFRTMAYDYKGGRAKTPDEFREQFFYIRQLQDVLGIKHYELEHYEADDII 120

LVAFDAGKTFFRTMAYDYK GRANTP+EPREQFFYIR+ L LGI +YELEHYEADDII

Sbjct: 61 LVAFDAGKTFFRTMAYDYKGRANTPEEFREQFFYIREMLTALGIAYELEHYEADDII 120

Query: 121 GTLAKQAEASNEHFDITVSGDKLQLQDNTNIVVEISKGVARFEPTPAYLMEKMGIT 180

GTL K AE + PD+VSGDKLQLQD NTIVVEISKGVARFEPTPAYLMEKMG+T

30 Sbjct: 121 GTLDKVAERTVFPDITVSGDKLQLQDNTNIVVEISKGVARFEPTPAYLMEKMGIT 180

Query: 181 PSQFIDLKALMGDKSDNIPGVTKIGERTGLKLLSEYSGLEGIYENIEMKQSKQKCNLIN 240

P+QFIDLKALMGDKSDNIPGVTKIGERTGLKLL E+GSGLEGIY+I+ K SKMKCNLIN

35 Sbjct: 181 PNOFIDLKALMGDKSDNIPGVTKIGERTGLKLLHEFGDSLEGIYEHIDGFKTSKQKCNLIN 240

Query: 241 DKFAFLSKTLATINTASPIITGLDILYSGPODKALQSQPYDEMDFQKQKALGEETSQ 300

D++QAFLSKTLATIN ASPITGL+DI+Y+GP D+ +LSQPYDEMDF Q K L + Q

Sbjct: 241 DRDQAFLSKTLATINTASPIITGLDIDIVNGP-DVASLSQPYDEMDFVQLKGLASQMRQ 299

40 Query: 301 EDFEV-DFTVEBQLKTEMFSDNDFYFEMLGDNVHVEDLGIAGNSDITIYATNSVLSQ 359

E V + EV + ++FS D +YFE L DNYH E +IG AMG+ + IYA+++ LL

Sbjct: 300 EPVAVISYQVTVNVSADLFSAEIDIFYFETLRDNYHREALIGFAMGHGBQIYASTDGLGLA 359

45 Query: 360 EALFKGALSFKPIKTYDFKRSKVLNKNFIDLPPEAPDTRIAKYLSTTDEDNVSTIARIY 419

FK+ KP+ TYDFKRSKVL++ I+L P++D RLA YLLST ENL +STIAR++

Sbjct: 360 TDSFKQVFKPIATYDFKRSKVLHSLGLTSLVARSVDARLANYLSTVEENELSTIARIF 419

Query: 420 TNLPLDITDAVYGKAKRAIPKTRFLEHLAKKVKVLVDSEANIMQQLKANQRELLFEM 479

T++ Ld +D VYGKAKRA+P+K LEHLA+KVKVL+DS++ ++ +L A+EQ +L

50 Sbjct: 420 TDISLEEDTVYGKAKRAVPDKVLEHLARKVKVLDDSKSMLDKT+AHSGQDLIYQNI 479

Query: 480 EQPIANVLAKMEIRGIRKVNKNLNMENAIENQKVIETITQEIYELAGQEFNINSKPKGLK 539

E PLANVLAKMEI GKRV + TL +MA +N+ +IE LTQEIY+AQGEFNINSKPKGLS +

55 Sbjct: 480 ELPIANVLAKMEISGIRKVNRAFLQDMABQNKVIIIEALTQEIYNAGQEFNINSKPKGLSI 539

Query: 540 LPETLGLVEMTKTKTKYSTAVVLERLAPI+P+V KIL+YRQTKLQSTY+IGLQYI 599

LFE + LP+ENTTKTKTKYSTAV+VLERLAPI+P+V KIL+YRQTKLQSTY+IGLQYI

Sbjct: 540 LFEKQQLPLEMTKTKTKYSTAVVLERLAPI+P+V KIL+YRQTKLQSTY+IGLQYI 599

60 Query: 600 LEDGKIHTRYVQDLTQGRILSSDPNLQNIPIRLEQQRILIRKAFVPSSEDNAVLSSSDYSQ 659

L DGIKHTRYVQDLTQGRILSS DPNLQNI+RLEQQRILIRKAF +S+AVILSSSDYSQ

Sbjct: 600 LADGKIHTRYVQDLTQGRILSSVDPNLQNIPIRLEQQRILIRKAFVPSHEDAVILSSSDYSQ 659

Query: 660 IELRVLAHISKDEHLIAAFEGADHTSTAMRVFGIEKPENVINDRRNAKAVNFGIVYG 719

IELRVLAHIS DEHLIAAF EGADHTSTAMRVFGI++ +VT NDRNRKAVNFGIVYG

65

-1270-

Sbjct: 660 IRKRVLAHISGEHLIAAFNRESADIHTSTAMRVFGIDRAADVIANDRNRKAKAVNFVYG 719
 Query: 720 ISDFGLSHNLGIPKRLAKQYIDTYPERYPGKKNVMEIVVREAKDKGYVETLPHRRSLSPD 779
 ISDFGLS+NLGI RK AK YIDTYPERYPGK YME VVREAKDKGYVETL RRR LPD
 Sbjct: 720 ISDFGLSHNLGITRKQAKSYIDTYPERYPGKAYMENNVREAKDKGYVETLPHRRSLSPD 779
 Query: 780 INSRNFNIQFAERTAINSPTQGSADILKIAMINLDRVLDGGYKSKMLLQVHDSIVLE 839
 INSRNFN+R FAERTAINSPTQGSADILKIAMINL+ L GG+++KMLLQVHDSIVLE
 Sbjct: 780 INSRNFNVRSFAERTAINSPTQGSADILKIAMINLKAQAGGFRAKMLLQVHDSIVLE 839
 Query: 840 VPNEETGAIRELATVTTMESATLSVPLADENAGETWYRAK 880
 VFN+E+ AI++IV TME+A+ L+VPL DE+ G+WYRAK
 Sbjct: 840 VPNEELTAIKKLVDTEMAAVDLAVPLCVDESTGHWSYRAK 880

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1135

A DNA sequence (GBSx1211) was identified in *S. agalactiae* <SEQ ID 3523> which encodes the amino acid sequence <SEQ ID 3524>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1880 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9571> which encodes amino acid sequence <SEQ ID 9572> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05860 GB:AP001514 unknown conserved protein [Bacillus halodurans]
 Identities = 72/134 (53%), Positives = 94/134 (69%), Gaps = 3/134 (2%)
 Query: 17 NPISDFMLKNYLTKAKTIAVVGLSGRQETAAYQVSKIMQEGYQIIPVNPKNAGQKILQK 76
 NPISD +K L +AK IAVVGLS + +Y VS MQ AGY+IIPVNP ++LG+
 Sbjct: 4 NPSEDEKIQQLQKAKIAVVGLSGNPQRTSYMSAAMHAGYETIIPVNP--TVDEVLGEK 61
 Query: 7 TYASLKDVTEHIDIVNIFRSEYLPDIAREFLEVDADIFWAQLGLESQRAETILKQAGHK 136
 SL+D+ +DIVN+FRSE+LPD+ARE +E+ A +FWAQLGLE++EA L+Q G
 Sbjct: 62 AVPSLQDLQEGADVIVNIFRSEHLPDVAARETVETGAPVFWAQLGLEKNKRAYDYLQQHGVT 121
 Query: 137 QIVNNKCLKVECEK 150
 I NN+C+KVE K
 Sbjct: 122 SI-MNRCKLVEHAK 134

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3525> which encodes the amino acid sequence <SEQ ID 3526>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0837 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/141 (61%), Positives = 114/141 (80%)

-1271-

Query: 11 MVYHFCNPSPDFMLKLYLTKAKTIAVVGLSDRQTAAYQVSKIMQEAQYLIIPVNPNGQ 70
 ++Y FCNPS+ +LK YL AKTIAVVGLSDR+TAY V+K MQ Y+IIPVNPK AGQ
 5 Subjct: 1 VIYSPFCNPSPDFMLKLYLTKAKTIAVVGLSDRKDTAAYGVAKPMQMDYRIIPVNPFLAQ 60

Query: 71 KILGMYVYASLKDVTEHIDVNIHRSRYLPDIARKEFLVDADIFWQALGLEQEAFTIL 130
 ILG+ YAS+K + +DIV+FRRSB+LP+AR+PL A +FWQALGLE+QEA+TIL
 Subjct: 61 LILGEKVYASIKAIPEVDIVDFRSEPLFEVARDPLASQAKVFWQALGLEWQEAQITL 120

10 Query: 131 KQAGHKQIVNNKCLKVCCQKL 151
 + AG + IVNN+CLK+ +L
 Subjct: 121 RSAGKEAIVNNRCLKIDVQL 141

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 1136

A DNA sequence (GBSx1212) was identified in *S. agalactiae* <SEQ ID 3527> which encodes the amino
 acid sequence <SEQ ID 3528>. Analysis of this protein sequence reveals the following:

Possible site: 13
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3367 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9573> which encodes amino acid sequence <SEQ ID 9574>
 was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3529> which encodes the amino acid
 30 sequence <SEQ ID 3530>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4960 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 113/151 (74%), Positives = 133/151 (87%), Gaps = 1/151 (0%)

Query: 7 MDSSHSHQRPIDAYENVLEHLREKRIRITETRKAIISYMNNSRHPSPAETIYNDLPEYP 66
 MD HSH + LDAYENVLEHLREK IRITETRKAIISYM+ S RHPSPA+KIY DL P +P
 Subjct: 1 MDIHS+QQALDAYENVLEHLREKRIRITETRKAIISYMIQSTHPSADKIYRDLQPF 59

45 Query: 67 NMSLATVYNNLKVLDGCFVTELELQNYSTTYDFMGHQHNLNACRDCGKIYDFVDVLL 126
 NMSLATVYNNLKVLDGCFV+ELK+ N TTYDFMGHQH+N+ CR CGKI DF+DVD++
 Subjct: 60 NMSLATVYNNLKVLDGCFVSELEKLNLTYYDFMGHQHVNVCETCGKIDFMDVDV 119

50 Query: 127 DISREAHQQTGFVETRVQLVAYGICPCRCRK 157
 DI++EAH+QTG++VTR+ ++AYGICP+QQ K
 Subjct: 120 DIAKEAHEQTGYKTRIFVIAYGICPCDQAK 150

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 55 vaccines or diagnostics.

-1272-

Example 1137

A DNA sequence (GBSx1213) was identified in *S. agalactiae* <SEQ ID 3531> which encodes the amino acid sequence <SEQ ID 3532>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.13    Transmembrane    16 - 32 ( 14 - 32)
      INTEGRAL    Likelihood = -1.81    Transmembrane    496 - 512 ( 496 - 515)

----- Final Results -----
bacterial membrane --- Certainty=0.1850 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:CAA06650 GB:AJ005645 ssrc [Staphylococcus aureus]
    Identities = 41/146 (28%), Positives = 63/146 (43%), Gaps = 13/146 (8%)

Query: 4  SQYNKNSIRRLKVGAAASVMIASGSIVALGQSHIVSAD---EMSQPKTITITAPTANTSTN 59
      ++ NK+SIR+ VG AS+++ + I L ++ E+Q K TAP+ N +T
20 Sbjct: 16 NRIKNSIRKYSQVGTASILVGTITLIFGLSGHEAKAHEHTINGELNQSKNETTAPSENKITT- 74

Query: 60 VESSTDKALSKVTITNETSSEMPK--MNNNAKVEKTSDEKMMVATSVRKQATPTPVAMT- 116
      D K T +++ PK M + A V++TS + T T T
25 Sbjct: 75 --KKVDSRQLKNTQTATADQPKVIMSDSATVKRTSSNMQSPQNATAMQSTTTKTSNVTIN 132

Query: 117 ---KTTTSVDEVKKSTDTAFKQTVDFP 139
      TT +E KS T K P
30 Sbjct: 133 DKSSTTYSNETKSNLTQAKDVSTTP 158

```

30 No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8735> and protein <SEQ ID 8736> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 7
MG: Discrim Score: -0.92
35 GVH: Signal Score (-7.5): -2.48
    Possible site: 39
    >>> Seems to have no N-terminal signal sequence
    ALOM program count: 2 value: -2.13 threshold: 0.0
      INTEGRAL    Likelihood = -2.13    Transmembrane    16 - 32 ( 14 - 32)
      INTEGRAL    Likelihood = -1.81    Transmembrane    496 - 512 ( 496 - 515)
      PERIPHERAL    Likelihood = 7.96    402
    modified ALOM score: 0.93

*** Reasoning Step: 3
45 ----- Final Results -----
      bacterial membrane --- Certainty=0.1850 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

50 LPXTG motif: 485-489

The protein has homology with the following sequences in the databases:

```

55 p|5981|5780 leukotoxin > Insert characterized
    SP|P16462|HLVA_ACTAC LEUKOTOXIN. > Edit characterized
    GP|I41834|gb|AA21922.1|M27399 leukotoxin (lta) [Actinobacillus actinomycetocombitans]
    Insert characterized

Query: 210 VSLNGNTTGKQALLDQI|AND---KHSYQATIRVYGAIDGKVLKRMISPMKVTINIP 266
      ++ NG+ + G+A +D +K + KHS + T ++ G +DL + +T P
60

```

-1273-

Sbjct: 488 ITRNGDRI-QSGKAYVDYLKKGELAKHSDEFTKQILDPIKGNIDLSGIKGSTTLTFLNP 546

Query: 267 HIITMEVKNSLMAFKKGV-DVPACKVSAKAKG-PPLAGVNE--TIPYEFGDGMLT 322
+ T E + + + E + + + K + K K G P G V + + A D +

Sbjct: 547 LLTAGKEERKTRQSGKYEFITELK/KGRTDMKVGVPNSNGVDFSNLIQHAVTRDNKVL 606

Query: 323 RLILKASBGAKNDSNGVDKNSPLL-----PLKDLTKGYFYQVSLNGNTAGKKGOALLD 376
L + GAK V S + + + D + KG + + + + G A K GQ + +

Sbjct: 607 EARLIANLGAKKDDYVFGVSGSTIVNAGDGYDVEDVYSKGRIG-ALITIDGRNATKAGQYKVB 665

Query: 377 QIKANGSHTYQATITIVYTKDGK 400
+ + G + Q T + + TK GKV

Sbjct: 666 R-DLGGTVLQRTVSKQSTKRGKV 688

SEQ ID 3532 (GBS1) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 3; MW 78kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 3; MW 53kDa).

The His-fusion protein was purified as shown in Figure 189, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1138

A DNA sequence (GBSx1214) was identified in *S.agalactiae* <SEQ ID 3533> which encodes the amino acid sequence <SEQ ID 3534>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3585 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB54578 GB:AJ006397 response regulator [Streptococcus pneumoniae]
Identities = 143/228 (62%), Positives = 183/228 (79%), Gaps = 1/228 (0%)

Query: 1 MTQKLLVDDRFELIIDINRRYLEQAGYEVSVAAADGIEALKEVDNRFDLIISDIMPYMD 60
N + LLVDD E I DI + RYL QAGY + V VA DG + EAL + D LII + D + M P + MD

Sbjct: 1 MGTITLLVDDVEITDIHQRYLIQAGYQVLVAHDGLEALELFKKPIDLIITDVMMPMD 60

Query: 61 GYDFISEVLVREPNOPLFETITAKVSEPKIYISLSMGADDFISKPPSPRELVLVRVKNILRR 120
GYD ISEV P QPFLFETAK SE DKYI LS + GADDFI + KPSPRELVLVRV NILRR

Sbjct: 61 GYDLISEVQLSPSPQPLFLITAKTSQDKITGLSLGADDFIAKPPSPRELVLVRVHNLRR 120

Query: 121 IYGNHQSEVILTIGDLVIDQKRLVMVDCNTISLHNKSFLLWILAMHLNVRVSKFELYE 180
++ + E + + + G + L + + V + + LT KS + LLWILR + + RVFSKT + LYE

Sbjct: 121 LR-RGGETELISLGNLKNHSSHEVQIGREMLDTVKSFEWLLASNPFRVPSKTDLYE 179

Query: 181 RVWGEFLDDTINTLNVIHALRDLAKFSDTNTPTIKTVWGLGYKLEE 228
++ N E + + + D D T I N L N V H I A L R + L A K + S + D T P T I K T V W G L G Y K + E +

Sbjct: 180 KINKEDYVDDTINTLNVIHALRDLAKFYSSDQPTTIKTIVWGLGYKIKK 227

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1274-

Example 1139

A DNA sequence (GBSx1215) was identified in *S. agalactiae* <SEQ ID 3535> which encodes the amino acid sequence <SEQ ID 3536>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.13    Transmembrane    42 - 58 ( 33 - 65)
      INTEGRAL    Likelihood = -7.54    Transmembrane    7 - 23 ( 3 - 29)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.4652 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
    >GP:CAB54579 GB:AJ006397 histidine kinase [Streptococcus pneumoniae]
    Identities = 190/343 (55%), Positives = 249/343 (72%)

    Query: 1   MKLKYYIVGYILSMILTIVAGVFFGLNMLIETRGVYYILSVTIIACIVGGIVNLFLLSS 60
    Sbjct: 1   MKLK YI++GY+IS L+T+ VF+ + MLI +Y++L +TI+A +VG ++LPLL
                MKLSYILVGYIISTLLTILVVFVAVQKMLLAKGEIYFILGMTIVASLVAGAGISLFLLLP 60

    Query: 61   VFTSLKLLKQMKDISQKCFDTKAQICSPOEFKDLTAFNQMSSELESTFKSLNESEREK 120
    Sbjct: 61   VFTSL KLK+ K++ + F+ ++ P EF+ L FN+MS +L+ +F SL ESEREK
                VFTSLGLKKEHAKRVAADFFPSNLEVQGPVEFQQIQTINEMSHDLQVSFDSLEBSEREK 120

    Query: 121  TMVIAQLSHDIKTFITSIQSTVEGHLGGIIEEEVNYLYLTISRQTNRLNHLVEELSFIT 180
    Sbjct: 121  GMVIAQLSHDIKTFITSIQATVEGHLGGIIESEQAHYLATIGRQTERLNKLVLEELNFLT 180

    Query: 181  LETMSDTAEPHKKEITYDKLLIDILSEFQLVFEKENRQVMIDVAPDVLSKQYDKLSR 240
    Sbjct: 181  LNTARNQVETTSKDSIFDKLLIDILSEFQFLTEGRDVRVHLQVIPESARIGSDYAKLSR 240

    Query: 241  ILINLISNAVYKSDPGSLTIKAYSNRQDIVIDIIDQGVGIKDEDLASIFNRLYRVSSR 300
    Sbjct: 241  ILVNLVNAFKYSAPQTKLEVVAKLEKDLQSLISVTEBQGIAPEDDENIFKRLYRVETSR 300

    Query: 301  NMKTGGHGLGLYIARQLABQLNGDILVESQYKQSKPSLVKL 343
    Sbjct: 301  NMKTGGHGLGL IAR+LAQLG +I V SQY GS P+LVL L
                NMKTGGHGLGLAIARELABQLNGGEITVSSQYGLSGTPTFLVNL 343
  
```

There is also homology to SEQ ID 1178.

A related GBS gene <SEQ ID 8737> and protein <SEQ ID 8738> were also identified. Analysis of this protein sequence reveals the following:

```

50  Lipop: Possible site: -1    Crend: 3
    MoG: Discrim Score:      8.67
    GVH: Signal score (-7.5): -5.75
    Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq
    ALOM program    count: 2 value: -9.13 threshold: 0.0
      INTEGRAL    Likelihood = -9.13    Transmembrane    42 - 58 ( 33 - 65)
      INTEGRAL    Likelihood = -7.54    Transmembrane    7 - 23 ( 3 - 29)
      PERIPHERAL Likelihood = 3.92      196

55  modified ALOM score: 2.33

    *** Reasoning Step: 3

    ----- Final Results -----
      bacterial membrane --- Certainty=0.4652 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```


-1276-

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -7.70 Transmembrane 125 - 141 (110 - 155)
 INTEGRAL Likelihood = -7.59 Transmembrane 38 - 54 (36 - 56)
 5 INTEGRAL Likelihood = -6.48 Transmembrane 146 - 162 (143 - 174)
 INTEGRAL Likelihood = -5.57 Transmembrane 72 - 88 (63 - 93)
 INTEGRAL Likelihood = -1.33 Transmembrane 229 - 245 (227 - 245)
 ----- Final Results -----
 10 bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9575> which encodes amino acid sequence <SEQ ID 9576>
 15 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA79984 GB:Z21972 ORF1 [Bacillus megaterium]
 Identities = 35/119 (29%), Positives = 62/119 (51%), Gaps = 15/119 (12%)
 20 Query: 142 SSFLLLSGNLLIAPLVIVSSLTITTKAVIKLV---QQYYSYSISTLVFTTQLESQNYEG 198
 +S++L+ +++ A + + S L+ +IK + QQ++ + YT LE+
 Sbjct: 105 TSPKLI-GRSILQAIPIFLMSLLILIPGIKAIAYSQQPFL--LWDEPETVLEA----- 156
 Query: 199 PSKLVASRRLANGNKLRLLSLPIGQWQPIATIFSPGLVIVYLLPYQTARLIPYKRI 257
 + S++ M G K + FL+ LSPIGN L +F+ G+ ++L+PY T FY +
 25 Sbjct: 157 ----ITRSKKRMKGLKWKYFIMHLSPIGWIQMPITIGILLWILPIAGTITTAAPYEL 211

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3539> which encodes the amino acid
 sequence <SEQ ID 3540>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.08 Transmembrane 148 - 164 (143 - 170)
 INTEGRAL Likelihood = -8.28 Transmembrane 114 - 130 (101 - 141)
 INTEGRAL Likelihood = -6.69 Transmembrane 60 - 76 (49 - 82)
 35 INTEGRAL Likelihood = -3.72 Transmembrane 21 - 37 (21 - 39)
 INTEGRAL Likelihood = -2.34 Transmembrane 222 - 238 (221 - 239)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA79984 GB:Z21972 ORF1 [Bacillus megaterium]
 Identities = 63/220 (28%), Positives = 100/220 (44%), Gaps = 31/220 (14%)
 45 Query: 62 LGLILSLFSLASPTMI-DVVRHPRQKVSFAESTTAPSKFPGLNLVLAITKMLFELIWS 120
 + L+L L+P++ F +I +V+ + T + F + +A+ L S
 Sbjct: 22 VSLMLLLFLINLVFPFLIVEVIGSGGPFSEWLMQSETPLWSIDIPMSVFSIALIP----LTIS 77
 50 Query: 121 LIWFF-----GLFIFLSGLSAFLVNAKSGSSSTVISLPLFGVLSLIGPGSI 167
 WFF I+ G +F + G+S + + L+ +L + G
 Sbjct: 78 TTFYFLNLVREGNPGIPSEVFAIYKDKGTSFKL---IGASILQAIPIFIMSLILLIPG--- 131
 55 Query: 168 YINRYAYSLSSEYLLYDEVKQGTLLGAIAVIETSVAMMKGYKWKLPFLQSLPTGWLFINI 227
 I + AYS +LL D R T L A I S MKG KWK F + LSP GW +L +
 Sbjct: 132 -IKAIAYSQQPFLKDH-PRYTVLEAIT--ESKKRMKGLKWKYFIMHLSPIGWILCM 186
 Query: 228 VTFGLNIYLLFPYPTTANVIFDYDLKKRPKKDD--PIEG 265
 T G+ ++L+PY T FY++L +D DD IEG
 60 Sbjct: 187 PTLIGILLWILPIAGTITTAAPYELIVPQEDIDDDQQTIG 226

-1278-

FLMHLSPFGWGLCMFTIAGIGLLWLIFPYAGTTTAAFYRELLVLPQRDIDDDQIQEG
 190 200 210 220

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1141

A DNA sequence (GBSx1217) was identified in *S. agalactiae* <SEQ ID 3541> which encodes the amino acid sequence <SEQ ID 3542>. This protein is predicted to be tRNA-guanine transglycosylase (tgt). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.3706 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9577> which encodes amino acid sequence <SEQ ID 9578> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:CB14731 GB:Z99118 tRNA-guanine transglycosylase [Bacillus subtilis]
 Identities = 269/377 (71%), Positives = 320/377 (84%)

Query: 12 MTDHPKIRYLKQEGHTGARLGEIITPHGTFFTPMPMPVGTQATVKTSPEELKEMSGSI 71
 M + P+Y IK+ K TGRLG++ THG+F TP+PMPVGT ATVKT SPEELK M +GI
 Sbjet: 1 MAEQPIRYEPIKECKOTGARLGKVTPHGSPETPVMPVGTLATVKTMSPEELKAMDAGI 60

Query: 72 ILSNTYHLWLRLPGDELIAGGLKPFNMMDQAILTDSGGFQVYSLADSNITREBVTFKN 131
 ILSNTYHLWLRLPG +++ +AGGLKPFNMMD+AILTDSGGFQV+SL+ RNI RECV F+N
 Sbjet: 61 ILSNTYHLWLRLPGQDIVKEAGGLKPFNMMDRAILTDSGGFQVFLSKFNRNIEBEGVHFRN 120

Query: 132 HILGAKMFLSPKKAISIQNNLGSIDMSFDECPQFYQFYDYVKKSISRTSRWAERGLNAH 191
 HILG K+FLSPKKA+ ION LGSDIM+FDCEP + YDY+K+S+ERTSRWAER INAH
 Sbjet: 121 HILGDKLFLSPKKAISIQNLGSDIMAFDCEPPYAEYDYMKRSVERTSRWAERCLNAH 180

Query: 192 RRPDQQLFGIVQGAGFEDLRQSAKDLSMDPFGYSIOGLVAGETHDMIAVLDTFVPM 251
 R +QQLFGIVQG +EDLR QSA+DL+S+DFPGY+IGGL+VGE D MI VL+FT P+
 Sbjet: 181 NRQDEQQLFGIVQGEYEDLRQSAKDLSMDPFGY+IGGLSVGEPIKDVMRVLEFTPL 240

Query: 252 LPMDKPRYLMGVGAPDSLIDAVIRGVMDPVCVLPTRIARNGTCMTSQGRVLVKNKAFED 311
 LP DKPRYLMGVG+PD+LID IRGVMDPVCVLPTRIARNGT T++GRL +KNKAF D
 Sbjet: 241 LPMDKPRYLMGVGAPDIALIDGAVIRGVMDPVCVLPTRIARNGTVFTAGRLNMIKNKAFED 300

Query: 312 FPLDNPDCDYCTKNYTRAYIRHLLKADRTFGIRLTSYHNLVFLNIMKDVQRALMDNGL 371
 F P+D CDYCTKNYTRAYIRHLL+ +ETFG+RLT+YHNL+FL++LM+ VEQAL +D L
 Sbjet: 301 FRPIDECDYCTKNYTRAYIRHLLRCNRTFGLRLTITYHNLFLHILMSQVRQALRDRGL 360

Query: 372 LEFRQDFMERYGYGMN 388
 +FR++F ERYGY N
 Sbjet: 361 GDFRERFFERYGYNKN 377

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3543> which encodes the amino acid sequence <SEQ ID 3544>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2590 (Affirmative)	< succ>
---------------------	-----	--------------------------------	---------

-1279-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 351/380 (92%), Positives = 368/380 (96%)

Query: 12 MTEDEPIKYRLIKQEKHTGARLGSEITPHGTFTTMMFPGTQATVKTQSPEELKEMSGGI 71
 MTD+PIKYRLIK EKHTGARLGSEITPHGTFTTMMFPGTQATVKTQSPEELK +GSGI
 Sbjct: 1 MTDYPIKYRLIKAEKHTGARLGSEITPHGTFTTMMFPGTQATVKTQSPEELKAI GSGI 60

Query: 72 ILSENTYHLWRFGDELIAGAGIHKFMNMDQAILTDSGGPQVYSLADSRNITEGVTFKN 131
 ILSENTYHILWRFGDELIAGAGIHKFMNMDQ ILTDSGGPQVYSLADSRNITEGVTFKN
 Sbjct: 61 ILSENTYHLWRFGDELIAGAGIHKFMNMDQAILTDSGGPQVYSLADSRNITEGVTFKN 120

Query: 132 HLNGKMFSLPEKAISIQNNLGSDDIMMSFDECPQFYQPYDVKKSISERTSRNAERGLAH 191
 HLNGKMFSLPEKAISIQNNLGSDDIMMSFDECPQFYQPYDVKKSISERTSRNAERGL AH
 Sbjct: 121 HLNGKMFSLPEKAISIQNNLGSDDIMMSFDECPQFYQPYDVKKSISERTSRNAERGLAH 180

Query: 192 RRFHDQGLFGIVQGAGFEDLRQSAADLVANDFPFGYSIGGLAVGESHENRVLDTFTPL 251
 RRFHDQGLFGIVQGAGFEDLRQSA DLV+MDFPFGYSIGGLAVGE+H+EMNAVLDFT P+
 Sbjct: 161 RRFHDQGLFGIVQGAGFEDLRQSAADLVANDFPFGYSIGGLAVGESHENRVLDTFTPL 240

Query: 252 LPNKPRYLMGVGAPDSLIDAVIRGVDMFDCVLPTRIAENGTMTS+GRLV+KNAKFAED 311
 LP +KPRYLMGVGAPDSLID VIRGVDMFDCVLPTRIAENGTMTS+GRLV+KNAKFAED
 Sbjct: 241 LPNKPRYLMGVGAPDSLIDGVIRGVDMFDCVLPTRIAENGTMTS+GRLV+KNAKFAED 300

Query: 312 FTFLDNCDCYTCQNYTRAVIRHLKKADETFGIRLTSYHNLVFLVNLKMQVQAINDNL 371
 FTFLD +CDCYTC+NY+RAYIRHLKKADETFGIRLTSYHNLVFLVNLKMQVQAINDNL
 Sbjct: 301 FTFLDNCDCYTCQNYTRAVIRHLKKADETFGIRLTSYHNLVFLVNLKMQVQAINDNL 360

Query: 372 LEFRQDFMERYGYGNENRNF 391
 LEFRQDF+ERYGY +NENF
 Sbjct: 361 LEFRQDFMERYGYGNENRNF 360

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1142

A DNA sequence (GBSx1218) was identified in *S. agalactiae* <SEQ ID 3545> which encodes the amino acid sequence <SEQ ID 3546>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2479 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9303> which encodes amino acid sequence <SEQ ID 9304> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10795> which encodes amino acid sequence <SEQ ID 10796> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16256 GB:Z99164 hypothetical protein [Schizosaccharomyces pombe]
 Identities = 42/91 (46%), Positives = 62/91 (67%), Gaps = 3/91 (3%)

Query: 6 FGIGLSSSRRCYHYHTKLDIVALKCAVQCKYVACYKCHDALERHCFAA-TKSDETFP-VL 63
 +G +D+ +RC+HYH+K D+VAL+C C+K+YAC++CHD L H F K+ P V+
 Sbjct: 13 YGKLVDNETRCFHYHAKADVVALRCGQCKRYVACFCQCHDELNTHFPLPWKAKHFIPCVI 72

-1280-

Query: 64 CGSCRQMLTLKEYK-TGFCPCYCRMLFNPNQ 93
 CG+C+ LT++EY+ T C YC FNP C+
 Sbjct: 73 CGACKNSLITVEEYRSTVCKYCNHFPNPKCK 103

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3547> which encodes the amino acid sequence <SEQ ID 3548>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2769(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 55/93 (59%), Positives = 62/93 (66%)

Query: 2 MQEYFGIGLSSSRICYHYHTKLDIVALKCAVQCKYACVKCHDALEEHCFPAATKSDETFF 61
 M + FGI LD RC HYHT LDIV LKCA QY YACV CHD L + H F T ET P
 Sbjct: 1 MTDGFGIDLEQYRCLHYHTPLDVLGLKCAQYTYACVCHDQLTNHAFTVPTGHQTSF 60

Query: 62 VLCSGRQMLTLKEYKTGFCPCYCRMLFNPNQ 94
 V+CG CR++L+ EY G CFCY+ FNP C R
 Sbjct: 61 VIQGHCRKLLSPRYGCGCCPCYQSPFNACHR 93

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1143

A DNA sequence (GBSx1219) was identified in *S.agalactiae* <SEQ ID 3549> which encodes the amino acid sequence <SEQ ID 3550>. This protein is predicted to be transport protein. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.45	Transmembrane	300 - 316 (292 - 321)
INTEGRAL	Likelihood = -1.17	Transmembrane	265 - 281 (265 - 281)

----- Final Results -----

bacterial membrane ---	Certainty=0.4779(Affirmative) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm ---	Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10113> which encodes amino acid sequence <SEQ ID 10114> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:NAF12002 GB:AE002075 transport protein, putative [Deinococcus radiodurans]
 Identities = 108/295 (36%), Positives = 174/295 (58%), Gaps = 4/295 (1%)

Query: 31 GAWINLVNNSQRESEQVADQFGIDIDDLRAPLDVEETSRISVEDDYTLVIVDVPTVEIRN 90
 G WI+ P+ EE +V+ G++D L+ PLD +E SR ED L+I+ +
 Sbjct: 21 GCWIDAAAPTEELARVSRITGLRDYLYKPLDPDERSRFRERDGLLIIMQTSYRLAED 80

Query: 91 NKSYSMTIPMGLIVTDNAVITTC-LBHLITLFDHFYRRRVKNFTFMKTRFVQLLYRMAE 149
 + Y T+P+GI+ TD+ +T C LE + V+ T K R QL RNA+
 Sbjct: 61 SDIPYDTVPLGLHTDHLVTVCSLEENPVVGVVSGLVRRVSTVKNRLTLQLFLRAQ 140

-1281-

Query: 150 LYLQALRTIDRQSKIEAQLSATRNLQIDMMLEKSIIVYLKASLKFNRIIVKLTST 209
 *L +R I+++ D IE ++EATRN +L+D+++LEKSYV LK NE +++++
 Sbjct: 141 RFLIDVRQINKRVDAIEDKMNATRNRELLDLLEKSLIVYPITGLKANRAMMERVKRDR 200

5 Query: 210 SSLKKYIEDEDLLETLIETQQAIEAMANIYENVLAMTETTASLIQNQNTIMKTALAVT 269
 + Y ED +LI+D LIE QAIRMA+I N+L +M AS+I NN N ++K L + T
 Sbjct: 201 I-FEMYEDSELDLVDLIENLQAIEAMASTANILTEMAGAPASVINNVNQVVKLVTTT 259

10 Query: 270 MTLIDPTVIPSAYGMNPQNMPLNGLANGPIYVVLAPLASSFVVFYPIRKKWF 324
 + + IPT++ +GMH + +P + +GP V+ +A +S + P F R K K F
 Sbjct: 260 ILVAIPTLVSGFFGMNVGG--LPSSDSPIGVLWMTVMAGIASLLAPLPIYRKKVF 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 715> which encodes the amino acid sequence <SEQ ID 716>. Analysis of this protein sequence reveals the following:

15 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.81 Transmembrane 293 - 309 (288 - 311)
 INTEGRAL Likelihood = -1.28 Transmembrane 255 - 271 (255 - 271)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4524 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 272/314 (86%), Positives = 296/314 (93%)

Query: 11 MKCMFLSTAIRFKKIEITPEPGAWINLVNPSQESQVADQPGIDIDLRAPLDVESTSRI 70
 30 Sbjct: 1 MKQMFLEAIRFKKIEITPEPGAWI LVNPSQES ++ADQF IDI DLRAPLDVESTSRI 60

Query: 71 SVEDDYTLIVDVPTIYERNNKSYMTIPLGIIVTNAVITTCLEHILTFDHPHRRVKN 130
 +VEDDYTL+IVDVP YBERNNKSY+T+P+GIIVT+NAVITTCLE +TFDHP+ RRVKN
 Sbjct: 61 AVEDDYTLIVDVPTIYERNNKSYMTIPLGIIVTNAVITTCLEHILTFDHPHRRVKN 120

35 Query: 131 FYTFMKTRFVFPQLYRNAELFLALRTIDRQSKIEAQLSATRNLQIDMMLEKSIIV 190
 FYTFMKTRFVFPQ-LYRNAEL-L ALRTIDRQ++EAQLS-ATRNE-LIDMMLEKSIIV
 Sbjct: 122 FYTFMKTRFVFPQI-LYRNAELFLALRTIDRQSERLEAQLS-ATRNE-LIDMMLEKSIIV 180

40 Query: 191 LKASLKFNRIIVKLTSTSSLKKYIEDEDLLETLIETQQAIEAMANIYENVLAMTETT 250
 LKASLKFNRIIVKLT+STSSLKKYIEDEDLLETLIETQQAIEA IYENVLAMTETT
 Sbjct: 181 LKASLKFNRIIVKLTSTSSLKKYIEDEDLLETLIETQQAIEAGIYENVLAMTETT 240

45 Query: 251 ASIIINNQNNTIMKTALAVTWLIDPTVIPSAYGMNPQNMPLNGLANGPIYVVLAPL 310
 ASII NNQNTIMKTAL+TM LIDPTVIPSAYGMNPQNM+PLNGL H P Y+ L+L+
 Sbjct: 241 ASIIINNQNNTIMKTALAVTWLIDPTVIPSAYGMNPQNMPLNGLAHAPWYITLAWLL 300

Query: 311 SSFVVYPIRKKWF 324
 SSFVV YPIRKKWF
 50 Sbjct: 301 SSFVVYPIRKKWF 314

SEQ ID 3550 (GBS257) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 3; MW 35kDa), in Figure 169 (lane 9 & 10; MW 50kDa) and in Figure 239 (lane 2; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 6; MW 60kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1282-

Example 1144

A DNA sequence (GBSx1220) was identified in *S.agalactiae* <SEQ ID 3551> which encodes the amino acid sequence <SEQ ID 3552>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -12.26    Transmembrane  158 - 174 ( 151 - 182)
INTEGRAL    Likelihood = -6.37     Transmembrane  93 - 109 ( 91 - 111)
INTEGRAL    Likelihood = -5.68     Transmembrane  188 - 204 ( 184 - 205)
INTEGRAL    Likelihood = -0.85     Transmembrane  118 - 134 ( 118 - 134)

----- Final Results -----
bacterial membrane --- Certainty=0.5904 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3553> which encodes the amino acid sequence <SEQ ID 3554>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -6.95     Transmembrane  92 - 108 ( 88 - 110)
INTEGRAL    Likelihood = -6.69     Transmembrane  153 - 169 ( 151 - 177)
INTEGRAL    Likelihood = -2.34     Transmembrane  183 - 199 ( 183 - 200)

----- Final Results -----
bacterial membrane --- Certainty=0.3781 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 135/217 (62%), Positives = 167/217 (76%), Gaps = 1/217 (0%)

Query: 1   MTLQDLTKKNQCFVHIATNQLLADGKSDAEIKAILLEHLPRIIDNQKGITARSILGAPT 60
          M LQ+LTKKNQCF+H ATN+L+ DGKSD +IK ILEE +P I+INCKKG+TAR+LLG PT
Sbjct: 1   MELQELTKKNQCFIHTATNKLIQDQKSDEDIKILEEAIPALEINQKGVITARNILGTP 60

Query: 61   TWAASPTERPEDKARVSVQKNTNFWLWMLDTSLLFLGLVTAINGLMLLFGQSNVNTGLIS 120
          WAASF+P KA KNFNFWLWMLDTSLLF+G+V LKNGM F + TGLIS
Sbjct: 61   AWAASFQDPSQA-AETDKNTNFWLWMLDTSLLFGIVALNGIMTFNTNATVGLIS 119

Query: 121  LITLGGGGASMYVTYYYYIYRHMGKPKSERPGWLKSFVADLVMLVWVWLFVAVVPLLEAT 180
          +L LGFGGGA+MY TTY+IYRH+GK KS RP W K A L+L ML+W AL++ LP +
Sbjct: 120  LLALGGGGASMYATYYTYFTHLGLKDKSLRPSWPKIIAALSGLMLWIALYSAFATPLTS 179

Query: 181  INPKLPEVVLFIILASFGRLRFYLQKYNIQSSNAFV 217
          +NP+LP + L II S LR+YLQKYNIQ+M+PV
Sbjct: 180  INPOLPFIALLIIGGVSLALRYLQKYNIQNTNSPV 216

```

A related GBS gene <SEQ ID 10787> and protein <SEQ ID 10788> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 8
McG: Discrim Score:      -9.94
GVH: Signal Score (-7.5): -3.66

Possible site: 29
>>> Seems to have no N-terminal signal sequence
ALOM program count: 4 value: -12.26 threshold: 0.0
INTEGRAL    Likelihood = -12.26    Transmembrane  158 - 174 ( 151 - 182)

```


-1283-

INTEGRAL Likelihood = -6.37 Transmembrane 93 - 109 (91 - 111)
 INTEGRAL Likelihood = -5.68 Transmembrane 188 - 204 (184 - 205)
 INTEGRAL Likelihood = -0.85 Transmembrane 118 - 134 (118 - 134)
 PERIPHERAL Likelihood = 8.43 50
 modified ALOM score: 2.95
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1145

A DNA sequence (GBSx1221) was identified in *S.agalactiae* <SEQ ID 3555> which encodes the amino acid sequence <SEQ ID 3556>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1349(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1146

A DNA sequence (GBSx1222) was identified in *S.agalactiae* <SEQ ID 3557> which encodes the amino acid sequence <SEQ ID 3558>. This protein is predicted to be excinuclease ABC (uvrA). Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10111> which encodes amino acid sequence <SEQ ID 10112> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC67271 GB:AF017113 excinuclease ABC subunit A [Bacillus subtilis]
 Identities = 642/940 (68%), Positives = 785/940 (83%), Gaps = 3/940 (0%)
 Query: 9 DKLMIRGAHNLKNIISVDIPDKLVVVITGLSGSGKSLAFTITVAGQRRYVESLSAYA 68
 D++ ++GARAHLKNI V IPRD+LVVVITGLSGSGKSLAFTITVAGQRRYVESLSAYA
 Sbjct: 4 DRIEVKGARAHNLKNIIDVTIPRQ+LVVVITGLSGSGKSLAFTITVAGQRRYVESLSAYA 63

-1284-

Query: 69 RQFLGNMSEKPDVDSIDGLSFAISIDQKTSKNPRSTVGTVEINDYLRLLYARVGTPYCI 128
 RQFLG M+KPDVD+I+GLSFAISIDQKTS+NPRSTVGTVEI DYRLLYARVG P+C
 Sbjct: 64 RQFLGQMCKPDVDIAEGLSFAISIDQKTSRNPSTVGTVEIYDYLRLLYARVGKPCP 123

5 Query: 129 NGHGAITASSVBOIVKVLALPERTFMOLAPIIRKRGQGHKSTFEKIQDGYVVRIDG 188
 IT++BQ+VD++L PERTK+Q+LAPI+ +KG H E++K GYVVRIDG
 Sbjct: 124 EHGIEITSQTIBQMVDRILEYFERTKQLVLAIPVSGRGGAHVKLBOIRKQGYVVRIDG 183

10 Query: 189 DIHDVTEVPKSEKSMHNDIVDRLINKGIRSELFDVSEAAALRLSDGVVVDITMDGNE 248
 ++ +++ EL K+K H+I+V+DR++ KEG+ +RL DS+E ALRL +G V+ID + E
 Sbjct: 184 EMALSDIDLEKKNKHSIEVVDIRIVVKEGVAARLSUSLETALRLGEGRVMDIVIGESE 243

15 Query: 249 LLFSEHYSCPGFPTVPELEPLRFSFNAPFGSCPTCDGLGKLEVDIDLVIPIRDSKITRE 308
 L+PSEH++CP GGF++ ELEPLRFSFN+PFG+CPCTDGLG+KLEVD DLVIP++ +L+E
 Sbjct: 244 LMFSEHACPHOGFSIGLEPLRFSFNPFPGACPTCDGLGKLEVDADLVIPIRDSKITRE 303

20 Query: 309 GALVPWNPFISSNYPIMLQNMVQFVDMDFPFKLSKABQDLALYSGSEREFHYHIND 368
 A+ PW PISS YYP +LE T+G+DMD P+ L K + D LYSGG+ +F Y ND
 Sbjct: 304 NAVAPWNPFISSQYPPQLLEAVCTHYGIDMDVFPKLEKHLQKLVLYSGGDIILYFRYND 363

25 Query: 369 FGGERNIDLFPFGVVRINRRYHETNSDYTNVMEYMNELKNCNCHGYRLNDQALCVRV 428
 FG R ++ FBGV+ NI RRY ET SD+ R M +YM++ C TC GYRL +AD V+
 Sbjct: 364 FQGIREFGIEQFGLVLRNIERYKETGSDFIREFGMEQVMSQKSCPTCKGYRLKKEALAVLI 423

30 Query: 429 GGEENIGQVSDLSIADHLELLETLRSSNEQLIARPIKEIHDRLSFLNIVGLNYLNL 488
 +G +IG+++LS+AD L + L LS + JA I++EI +RLSFL+ VGL+V+L L
 Sbjct: 424 ---DGRHIGKITELSVADALAFFKDLTISEKMDQIANLIREIVERLSFLDKVGLDYTL 480

35 Query: 489 SRAGTLGGGSEQRIRLATQIGSNLGLVYLDEPSIGLHQDRNDRLIDSLKKMRDLNT 548
 SR+AGTLGGGE+QRIRLATQIGS LSGVLY+LDEPSIGLHQDRNDRLI +LK MRDLNT
 Sbjct: 481 SRAGTLGGGSEQRIRLATQIGSRLGLVYLDEPSIGLHQDRNDRLISALKMRDLNT 540

40 Query: 549 LIVVEHDEDTMMAADNLIDVGGAGAGFGEIVASSTPKQVAKNTKITGQYLSGKKVIP 608
 LIVVEHDEDTMMAAD+LID+GGAG GG+++++GTP++V ++ S+TG YLSGKK IP+
 Sbjct: 541 LIVVEHDEDTMMAADYLIDIGGGAGIHGGQVISAGTPEEVVEDPNSLTGYSLSGKKPIPL 600

45 Query: 609 PSERRVGNRFLFEIKGAENNLQNLDVVKFLGKFIATVGVSGSGKSTLINSILKKAQAQ 668
 P ERR +GR++EIKGA+ENNL++ ++ KFLG F AVTGVSGSGKSTL+N IL KA+AQK
 Sbjct: 601 PERRKPDGRYIEIKGASENNLKKVNAKFLPQTFTAVTGVSGSGKSTLVNELIKKAQAQ 660

50 Query: 669 LNRNSDKPGKYVSLGIEYVERLIDQSPIGRTFRSNPATYTGVEDDIRDLFAQTNEAK 728
 L++ KPG + ++G++++D++IDIDQ+PIGRTFRSNPATYTGVEDIRD+FAQTNEAK
 Sbjct: 661 LHKAKAPGSHKETKGLDHLKVIDIDQAPIGRTFRSNPATYTGVEDIRDVFAQTNEAK 720

55 Query: 729 IRGYKGRFSFNVKGRCECSQGDGIKIEHMFLEPDVYVPCVCHGTRYNSSETLEHYKE 788
 +RGYKGRFSFNVKGRCE+C GDGIKIEHMFLEPDVYVPCVCHG RYN ETELY YK
 Sbjct: 721 VRGYKGRFSFNVKGRCEACGDGIKIEHMFLEPDVYVPCVCHGKRYNRETELYTYG 780

60 Query: 789 KNLAIQILDVTVNDVTFPAALFKIARKLQTIKDVLGVYVTLGQPATILSGEAGKRLAS 848
 K+I+ +LDMTV DA++FF IKI RLQCT+ DVLGLY+TLGQPATILSGEAGKRLAS
 Sbjct: 781 KSIIDVLMTVEDALSFPFNIPKIKRKLQTYDVLGYVTLGQPATILSGEAGKRLAS 840

Query: 849 ELHKRSTGKSLYILDEPTTGLHADDIARLLKVLDRFVDDGNTVLVIEHNLDVIRKADHII 908
 ELHKRSTG++LYILDEPTTGLH DDIAKLL VL R VD+G+TVLVIEHNLD+IKTAD+I+
 Sbjct: 841 ELHKRSTGRTLYILDEPTTGLHVDLDARLLVVLQRLVNDGDTVLVIEHNLDIIRKADYIV 900

Query: 909 DLGPEGGGGGGQIVAGTPEEVAANPKSYTOYLLKEKLAR 948
 DLGPEGG GGG IVA GTPPE+ R +SYTG YLK + R
 Sbjct: 901 DLGPEGGAGGGTIVASGTPPEITVSESTYGRYLKPVIER 940

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3559> which encodes the amino acid sequence <SEQ ID 3560>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1285-

bacterial cytoplasm --- Certainty=0.1138 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 835/940 (88%), Positives = 896/940 (94%)

Query: 7 NQDKIATIRGARAHLNLIQNSVDIFRDLKLVVVTGLSGSGKSLAFDTIYABQRRYVESLSA 66
 NQ-K+I GARAHLNLI V+I PRDLKLVVVTGLSGSGKSLAFDTIYABQRRYVESLSA
 10 Sbjct: 11 NQDKIITHGARAHLNLIQNSVDIFRDLKLVVVTGLSGSGKSLAFDTIYABQRRYVESLSA 70

Query: 67 YARQFLANMEKFDVDSIDGLSPAISIDQKTSKNRSTVGTVTEINDYLRLLYAVGTFF 126
 YARQFLANMEKFDVDSIDGLSPAISIDQKTSKNRSTVGTVTEINDYLRLLYAVGTFF
 15 Sbjct: 71 YARQFLANMEKFDVDSIDGLSPAISIDQKTSKNRSTVGTVTEINDYLRLLYAVGTFF 130

Query: 127 CINGHGAITASSVEQIVDKVLALPERTMQLIAPIRRKQGHKSTFEKLQKDGVRVVR 186
 CINGHGAITASS EQIV+VVALPERT+MQLIAP+RRKQGHK FEKIQKDGVRVVR+
 Sbjct: 131 CINGHGAITASSAEQIVQVVLALPERTMQLIAPVRRKQGHKSTFEKIQKDGVRVVR 190

Query: 187 DGDIDHDTVEFELSKSKHNIIDIVDRLINKEGIRSELPSVEAALRLSGDGVVIDTMG 246
 DGDIDHDTVEFELSKSKHNI+V+DRL+NK+GIRSELPSVEAALRL DGY+I+IDTMG
 Sbjct: 191 DGDIDHDTVEFELSKSKHNIIEVDRLINKEGIRSELPSVEAALRLSGDGVVIDTMG 250

Query: 247 NELLFSEHYSCPVCGFTVPELEPRFLSPNAPFGSCPTCDGLGIKLEVDIDLVIPIRDKSL 306
 NELLFSEHYSCP CGFTVPELEPRFLSPNAPFGSCPTCDGLGIKLEVD+DLV+D SK+L
 25 Sbjct: 251 NELLFSEHYSCPVCGFTVPELEPRFLSPNAPFGSCPTCDGLGIKLEVDLVLVDPDSKSL 310

Query: 307 REGALFVNPISSNYPTMLEQANTPGVDMOTPFPEKLSKAEQDLALYSGGERPFHFHY 366
 REGAL FVNPISSNYPTMLEQAN PGVDMOTPF L+ +E+DL YSGG+REFHFHY+
 30 Sbjct: 311 REGALFVNPISSNYPTMLEQANAPGVDMOTPFPEALTEERDLVLVSGGERPFHFHY 370

Query: 367 NDPGGERNIDLPFGVVMNINRRYHETNSDYTRNVMRETNELKNTCHGYRLMDQALCV 426
 NDPGGERNID+PFGGV N+NRRYHETNSDYTRNVMR YHREL C TCHGYRLMDQALCV
 35 Sbjct: 371 NDPGGERNIDIPFGVVTNVRNRYHETNSDYTRNVMRGYTMELTCTCHGYRLMDQALCV 430

Query: 427 RVGGEGLNIGQVSDLSADHLELETLRLSGNEQLIARPIKEIHDLRLPLNNVGLAYL 486
 VVGEGG +IGQ+S+LSADH+LLE L L+ NE IA+PI+KEIHDLRLPLNNVGLAYL
 Sbjct: 431 RVGGEGLNIGQVSDLSADHLELETLRLSGNEQLIARPIKEIHDLRLPLNNVGLAYL 490

Query: 487 NLSRAAGTSLGGESQRIRLATQIGNSLGVLYLDEPSIGLHQRDNRLIDSLKKRDLG 546
 LSR+AGTSLGGESQRIRLATQIGNSLGVLY+LDEPSIGLHQRDNRLI+SLKKRDLG
 40 Sbjct: 491 NLSRAAGTSLGGESQRIRLATQIGNSLGVLYLDEPSIGLHQRDNRLIDSLKKRDLG 550

Query: 547 NTLIVVEHDEDTMADWLIDVGPAGAGPGEIVASGTPKQVAKNTKISITGOYLSSKKVI 606
 NTLIVVEHDEDTM ADWLIDVGPAGAG PGEI ASGTPKQVAKN KSIITGOYLSSKKI
 45 Sbjct: 551 NTLIVVEHDEDTMADWLIDVGPAGAGPGEITASGTPKQVAKNKISITGOYLSSKKFI 610

Query: 607 PVPSERRVGNRFLIEIKGAENNLQNLVVKPLGKFIATVGVSGSGKSTLINSILKKA 666
 PVP ERR GNRF+EIKGAA+NNLQ+LDV+PPLGKFIATVGVSGSGKST+NSILKKA
 50 Sbjct: 611 PVPSERRVGNRFLIEIKGAENNLQNLVVKPLGKFIATVGVSGSGKSTLINSILKKA 670

Query: 667 QKLNKNSDKPGKYVSLGIEIYVDRLLIDQSPIGRTPRSNPATYITGVDDIRDLFAQ 726
 QKLNK+DKPGKY S+ GIE++RLIDIDQSPIGRTPRSNPATYITGVDDIRDLFAQ
 55 Sbjct: 671 QKLNKNSDKPGKYVSLGIEIYVDRLLIDQSPIGRTPRSNPATYITGVDDIRDLFAQ 730

Query: 727 AKIRGYKGRFSPFNVGKGRCEACSGDGIKTEHMFLDVYVPCVCHGRYNSSETLEVHY 786
 AKIRGYKGRFSPFNVGKGRCE+CSGDGIKTEHMFLDVYVPCVCHG RYNSSETLEVHY
 60 Sbjct: 731 AKIRGYKGRFSPFNVGKGRCEACSGDGIKTEHMFLDVYVPCVCHGRYNSSETLEVHY 790

Query: 787 KKNIAQLDMTVDAVTFPAI FKIDARKLQTIKDVGLGYTLGQPATTLSGGEAQRKML 846
 K KNIA+LDMTV+DA+ FP+AIPIARK+QTIKDVGLGYTLGQPATTLSGGEAQRKML
 65 Sbjct: 791 KKNIAEVLDMTVDAVTFPAI FKIDARKLQTIKDVGLGYTLGQPATTLSGGEAQRKML 850

Query: 847 ASELHKGSTGKSLYILDEPTTGLHADIARLLKVLDRFVDGNTVVLVIEHNLVDIKTADH 906
 ASELHKGSTGKSLYILDEPTTGLH DDIARLLKVL+RFVDGNTVVLVIEHNLVDIK+ADH
 Sbjct: 851 ASELHKGSTGKSLYILDEPTTGLHADIARLLKVLDRFVDGNTVVLVIEHNLVDIKSADR 910

-1286-

Query: 907 IIDLGPEGGIGGGQIVAIQTPEEVAHPKSYTGYYLKEKL 946
 IIDLGPEGG GGGQIVA GTPEEVA+ +SYTG+YLK KL
 Sbjct: 911 IIDLGPEGGDGGGQIVATGTPEEVAQVKESYTGHYLKEKL 950

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1147

A DNA sequence (GBSx1223) was identified in *S. agalactiae* <SEQ ID 3561> which encodes the amino acid sequence <SEQ ID 3562>. Analysis of this protein sequence reveals the following:

10 Possible site: 60
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.40	Transmembrane	471 - 487 (463 - 490)
INTEGRAL	Likelihood = -9.29	Transmembrane	246 - 262 (242 - 264)
INTEGRAL	Likelihood = -7.27	Transmembrane	183 - 199 (178 - 207)
INTEGRAL	Likelihood = -5.41	Transmembrane	351 - 367 (349 - 370)
INTEGRAL	Likelihood = -4.41	Transmembrane	87 - 103 (83 - 107)
INTEGRAL	Likelihood = -3.24	Transmembrane	375 - 391 (374 - 392)
INTEGRAL	Likelihood = -2.97	Transmembrane	17 - 33 (16 - 35)
INTEGRAL	Likelihood = -2.28	Transmembrane	420 - 436 (420 - 438)
INTEGRAL	Likelihood = -1.97	Transmembrane	320 - 336 (320 - 337)
INTEGRAL	Likelihood = -1.75	Transmembrane	214 - 230 (214 - 230)
INTEGRAL	Likelihood = -1.75	Transmembrane	288 - 304 (288 - 304)
INTEGRAL	Likelihood = -1.70	Transmembrane	110 - 126 (110 - 126)
INTEGRAL	Likelihood = -0.69	Transmembrane	152 - 168 (151 - 168)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB21192 GB:Z99106 similar to multidrug resistance protein [Bacillus subtilis]
 Identities = 198/481 (41%), Positives = 300/481 (62%), Gaps = 24/481 (4%)

35 Query: 9 IHGKPNYNTAMITILLIATPAQVLNQTSLTAIPTLMSFNHSLSTAQAQATTWFLANGI 68
 I KP+NR+ ++ +LL F +LNQT L TA+P +M FN+ + AQ TT F+L NGI
 Sbjct: 5 IEQKPFNRSVIVGILLGAFVAILNQTLLITALEPHINRDPFVDANQAGWLTTSFMLNNGI 64

40 Query: 69 MIPVSAYLATRFSTKLVVTSYVVLLIGLLMTLAPTSHNHLFWGRIQAISVGISMPL 128
 +IP+AA+L +F+++ L +T+ + G ++ AP N+ + L RIIQA GI MPL
 Sbjct: 65 LIPITAFLEIKFTSRALLITAMSIPTAGTVVGAFAP--NFFVLITARIQAAGAGIMPL 122

Query: 129 MOVVMNVFPBPORGANGKLVVGLAPAIGPILAGWILKQSPHFAHGLDITWRAIFLE 188
 MQ V + +PP E+RG AMG+ GLV+ APAIGPIL+GW ++ +NR+FF +
 45 Sbjct: 123 MQTVFLTIPIPEKRGQMGVGLVISFAPAIGPILSGMAVEA-----FNRSLFYII 174

Query: 189 LLILITVITILSPFVLKDVINKSVKLEVPSLILSLIIGFSGPLMGPTNVATYQWDIGYVI 248

-1287-

Query: 429 LSSVAQNIITNNKPSKDLLTMNPLKYANQMLNASLDGFHVSFAIGFVFAVLGLLWSLPLRK 489
 L SV N + + +A+L G + +P + V A++G I+S I+K
 Sbjct: 414 LVSVMNQNAH-----AGTTNVKHAALHCHMAAFIVAANIALVGFILSLPTLRK 461

5 There is also homology to SEQ ID 46.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1148

10 A DNA sequence (GBSx1224) was identified in *S.agalactiae* <SEQ ID 3563> which encodes the amino acid sequence <SEQ ID 3564>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.81 Transmembrane 8 - 24 (5 - 30)
 INTEGRAL Likelihood = -7.32 Transmembrane 36 - 52 (31 - 54)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 20

A related GBS nucleic acid sequence <SEQ ID 10109> which encodes amino acid sequence <SEQ ID 10110> was also identified.

A related GBS gene <SEQ ID 8743> and protein <SEQ ID 8744> were also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 9.52
 GVH: Signal Score (-7.5): -3.4
 Possible site: 22
 >>> Seems to have an uncleavable N-term signal seq
 30 ALOM program count: 1 value: -7.32 threshold: 0.0
 INTEGRAL Likelihood = -7.32 Transmembrane 11 - 27 (6 - 29)
 PERIPHERAL Likelihood = 11.19 130
 modified ALOM score: 1.96
 35 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 SEQ ID 8744 (GBS29) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 2; MW 25.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 6; MW 51kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1149

A DNA sequence (GBSx1225) was identified in *S.agalactiae* <SEQ ID 3565> which encodes the amino acid sequence <SEQ ID 3566>. This protein is predicted to be aminopeptidase P (pepQ). Analysis of this protein sequence reveals the following:

```

5      Possible site: 41
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.0724(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAA70068 GB:Y08842 aminopeptidase P [Lactococcus lactis]
      Identities = 44/126 (34%), Positives = 78/126 (60%)

      Query: 6  RLTRQQTALSQLSCDALILTNLTNIFYLTFGSGINATVLISPKHRIFVTDISKYIALIAQNT 65
      R+ + + + + D+LLIT+ NIFYLTFGSGT TV ++ K IF+TDSKI+ +A+
20      Sbjct: 2  RTEKLKVKMLTENIDSLITDMKNIFYLTFGSGTAGTVFLTQKRNIFMTDSKYSENAGRL 61

      Query: 66  VREFDIIISREPLAAILKIIDDALIAIGFETDISYHMYKHMVEVFEDYLRLEAPSVVEK 125
      ++ F+II +R+P++ + + + + + FE + Y +K + + L + V +
25      Sbjct: 62  IKNFIIIEIRDPISLLITELASGESVKNMAFEETVDYAFFKRLSKAATKLDLFSTNSNFVLE 121

      Query: 126 LRMIKD 131
      LR IKD
30      Sbjct: 122 LRQIKD 127

```

There is also homology to SEQ ID 3568.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1150

A DNA sequence (GBSx1226) was identified in *S.agalactiae* <SEQ ID 3569> which encodes the amino acid sequence <SEQ ID 3570>. This protein is predicted to be aminopeptidase P (pepQ-2). Analysis of this protein sequence reveals the following:

```

35      Possible site: 44
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.2508(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45      >GP:CAA70068 GB:Y08842 aminopeptidase P [Lactococcus lactis]
      Identities = 131/205 (63%), Positives = 163/205 (78%), Gaps = 3/205 (1%)

      Query: 2  LDFIKPDRITTELQVANFLDFRMRELGATGSPSDFIVASGYRSAMPHGVASQKTIQSGEITL 61
      L FI+P RT E+VANFLDF+NR+L A+G SF+ IVASG RS++PHGVA+ K IQ G+ +
50      Sbjct: 149 LRPIEFGRIT-EIEVANFLDFRMRLGASGISFETIVASGRKSSLPHEGVATSKNIQFGDPV 207

      Query: 62  TIDFGCYQHIVSDWTRTFTHIGHVTDQRKHIDIVLKSNAQIIGNVKSQMRCDYDYLAR 121
      T+DFGCFY+HY SDWTRTI +G V D+ R IY+ V K+N+A+I VK+GM YD + R
55      Sbjct: 208 TIDFGCYQHIVSDWTRTFIVGVSDDKMRITTYSTVRKANEALIKQVKAQMTYQXDNIPR 267

      Query: 122 QVINSYGYGNHPTHGIGHGMGLDVHRIPIYFGKS--EGVIASGMVVTDEPGIYLDNKYQVR 179

```

-1289-

+VIE + +G +FTHGIGHG+GLDVHEIPYF +S E + SGMV+TDEPGIYL GVR
 Sbjct: 268 EVIEKADFGQYFTHGIHGHLGLDVHEIPYFNQSMITENQLRSGMWITDEPGIYLPFGVGVR 327

Query: 180 IEDDLITETGCEVLTSAPKRLIVL 204
 IEDDLL+TE GCEVLT APKRLIV+
 Sbjct: 328 IEDDLVTENGCEVLTSAPKRLIVI 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3567> which encodes the amino acid sequence <SEQ ID 3568>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1450 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/203 (71%), Positives = 171/203 (83%)

Query: 2 LDFIKPRTTELQVANFLDFPMRELGAQPSDFDIVASGYRSAMPNGVASQKTIQSGETL 61
 LDFIKP TTE +ANFLDFPMR+ GA+G SFD IVASGY SAMPNG AS K IQ+ E+L
 Sbjct: 168 LDFIKPRTTTERDLANFLDFPMRQYGASGTSFDIIVASGYLSAMPNGRASKVQNKRL 227

Query: 62 TLDPGCYQHYVSDWMTRIHIGVTDQEREIYDIVLKSQAIIGNVKSGMKRCDYDLAR 121
 T+DFGCY HYVSDWMTRIHIG VTD+EREIY +VL +N+A+I +GM D+D + R
 Sbjct: 228 TMDFGCYNHVSDWMTRIHIGQVTDEREIYALVLAANKALAKASAGMTYSDFDGPR 287

Query: 122 QVIENSQYGRHFTHGIGHGMLDVHEIPYFGKSEGVIASGMVVITDEPGIYLDNKYGVRIE 181
 Q+I +GYG+ FTHGIGHG+GLD+HE P+PKSE ++ +GMVITDEPGIYLDNKYGVRIE
 Sbjct: 288 QLITEVQYGRHFTHGIGHGILDHENPFFGKSEQLLQAGMVVITDEPGIYLDNKYGVRIE 347

Query: 182 DDLITETGCEVLTSAPKRLIVL 204
 DDL+IT+TGC+VLT APKRLIVL
 Sbjct: 348 DDLVITTTGCVLTAPKRLIVL 370

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1151

A DNA sequence (GBSx1227) was identified in *S.agalactiae* <SEQ ID 3571> which encodes the amino acid sequence <SEQ ID 3572>. This protein is predicted to be yfhC protein (comEB). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1401 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:HAB05053 GB:AP001511 late competence operon required for DNA
 binding and uptake [Bacillus halodurans]
 Identities = 78/146 (53%), Positives = 107/146 (72%)

Query: 1 MNRLSWEDYFMANRLISKRSTCDKAFVGAULVKNRRIATGYNGGVBTDONNEVGHY 60
 MNR+SW+ YFMA +L++ RSTC R VGA +V++ RIITA GYNG +S +C + G Y+
 Sbjct: 1 MNRLSWDQYFMAQSHLLALRSTCTRLMVGATTVDRKRITAGGNGYSISGPHCIDEGCY 60

Query: 61 EDGHCIRTVHAEMNALIQCAKEGISTNNTIYVTHFFPCINCTKALLQAGVKIKITYKANYR 120
 +GHCIRTVHAEMNALIQCAK G+ T EIYVTHFFPC+NCTKA++Q+G+KK+ Y +Y+
 Shjct: 61 VEGHCIRTVHAEMNALIQCAKFGVPTGARIYVTHFFPCVNCCTKAIQSGIKKYYATDYK 120

Query: 121 HPFAIELMEAKGVAYVQHDVPTLG 146
 P+A EL GV Q ++ E+ L
 Shjct: 121 NSPYARELFRDAGVDVEQVELEML 146

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3573> which encodes the amino acid sequence <SEQ ID 3574>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3155 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/146 (91%), Positives = 140/146 (95%)

Query: 2 NRLSWEDYFMANAELISKSTCDRAFGAVLVKNNRIIATGYNGGVSETDNCNEGHYME 61
 NRLSW+DYFMANAELISKSTCDRAFGAVLVK+NNRIIATGYNGGVSE TDNCNE GHYME
 Shjct: 16 NRLSWQDYFMANAELISKSTCDRAFGAVLVKNNRIIATGYNGGVSEATDNCNEAGHYME 77

Query: 62 DGHCI RTVHAEMNALIQCAKEGISTNNTIYVTHFFPCINCTKALLQAGVKIKITYKANYRP 121
 DGHCI RTVHAEMNALIQCAKEGIST+ TEIYVTHFFPCINCTKALLQAG+ KITYKA+YRP
 Shjct: 76 DGHCI RTVHAEMNALIQCAKEGISTDTEIYVTHFFPCINCTKALLQAGITKITYKAHYRP 137

Query: 122 HPFAIELMEAKGVAYVQHDVPTLG 147
 HPFAIELME KGVAYVQHDV++ LG
 Shjct: 138 HPFAIELMEKGVAYVQHDVPTILG 163

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1152

A DNA sequence (GBSx1228) was identified in *S.galactiae* <SEQ ID 3575> which encodes the amino acid sequence <SEQ ID 3576>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2454 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1291-

Example 1153

A DNA sequence (GBSx1229) was identified in *S.agalactiae* <SEQ ID 3577> which encodes the amino acid sequence <SEQ ID 3578>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.65    Transmembrane    4 - 20 ( 3 - 21)

----- Final Results -----
      bacterial membrane --- Certainty=0.1659 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1154

A DNA sequence (GBSx1230) was identified in *S.agalactiae* <SEQ ID 3579> which encodes the amino acid sequence <SEQ ID 3580>. Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA04699 GB:AP001510 unknown conserved protein [Bacillus halodurans]
Identities = 47/94 (50%), Positives = 65/94 (69%)

Query: 2  LLFVGSVYVLIDGNQKLVINRGAIVBQBGQVYFYDYLQGIIPBGLINLQVYFYFQDID 61
      +LP+GS+VYL +G KL+I+NRG I+E G+ FDY G +P+GL ++V+YFM E+ID
Sbjct: 1  MLPTGSIVLKLKGTSLKMLINRGFILEANGENOMFDYSGCFYFQSLVPDKVFFYFHNID 60

Query: 62  EVVTEGYHDEEHEERSRLIEKWNTEGKNLPKKG 95
      EVVTEG+ D+EE+R +L WK KKK
Sbjct: 61  EVVTEGQDDEBQRFQKLPHDMCKENKIDRYVKKG 94

```

- 40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1155

- 45 A DNA sequence (GBSx1231) was identified in *S.agalactiae* <SEQ ID 3581> which encodes the amino acid sequence <SEQ ID 3582>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3560 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-1292-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1156

A DNA sequence (GBSx1232) was identified in *S.agalactiae* <SEQ ID 3583> which encodes the amino acid sequence <SEQ ID 3584>. This protein is predicted to be elongation factor p (efp). Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3067(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14376 GB:Z99116 elongation factor P [Bacillus subtilis]
Identities = 89/186 (47%), Positives = 120/186 (63%), Gaps = 1/186 (0%)

Query: 1 MIEASKLKAGMTFETADGKLIRVLEASHHKPGKNTIMMKLRDVRTGSTFDTSTYRPEEK 60
MI + + G+T + DG + RV++ H KPGKG +R KLR++RTG+ + ++R EK
Sbjct: 1 MISVNDPFRYGLTIDV-DGGIMRVVDFQHVKPGKGAFAFVSKLRNLKGTGAIQKRTFRAGEK 59

Query: 61 FEQAIIETVPAQYLYKMDTAYFMNNETDYQYEIFTVNIENSLLYILENSEVKIQPYGTE 120
+A IET QYLY D FM+ +Y+Q E+ IE EL Y+LEN V I Y E
Sbjct: 60 VAKAQIETKIMQILYANGQHVFMUTSSSYEQLELSATQIREELKYLLNENSVHIMVYQDE 119

Query: 121 VIGVQIFPTTVELTVAETQPSIKGATVTGSGKPAIMETGLVNVNPDFIEAGQKLIVINTAEG 180
+G+++P TVEL V ET+P IKG T +G KPA RTGLVNVNP F+ G LV+NT+G
Sbjct: 120 TLGIELFNTVELKVVETEPQIKGDTASGGTKPAKTETGLVNVNPPFVNEGDTLVNTSDG 179

Query: 181 TVYSRA 186
+TVSRA
Sbjct: 180 STYSRA 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3585> which encodes the amino acid sequence <SEQ ID 3586>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1813(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/186 (91%), Positives = 180/186 (96%), Gaps = 1/186 (0%)

Query: 1 MIEASKLKAGMTFETADGKLIRVLEASHHKPGKNTIMMKLRDVRTGSTFDTSTYRPEEK 60
MIEASKLKAGMTFE A+GKLIRVLEASHHKPGKNTIMMKLRDVRTGSTFDT+YRP+EK
Sbjct: 1 MIEASKLKAGMTFE-ASGKLIRVLEASHHKPGKNTIMMKLRDVRTGSTFDTTITRPEEK 59

Query: 61 FEQAIIETVPAQYLYKMDTAYFMNNETDYQYEIFTVNIENSLLYILENSEVKIQPYGTE 120

```

      FEQALISTVPAQYLYKMDQYAYFMN +TYDQYELP N+E ELLYILENS+VKIQFYQG+N
      Sbjct: 60 FEQALISTVPAQYLYKMDQYAYFMN+TYDQYELPVANVEQELLYILENSDVKIQFYQGE 119

5  Query: 121 VIGVQIVPTTVELTVAETOPSIKGATVIGSGKPAIMETGLVNVNVEFDIAGOKL+INTARG 180
      VIGV +PTTVELTVAETOPSIKGATVIGSGKPAI+ETGLVNVNVEFDIAGOKL+INTARG
      Sbjct: 120 VIGVTVFTTVELTVAETOPSIKGATVIGSGKPAIETGLVNVNVEFDIAGOKL+INTARG 179

      Query: 181 TYVSRA 186
      TYVSRA
10  Sbjct: 180 TYVSRA 185

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1157

- 15 A DNA sequence (GBSx1233) was identified in *S. agalactiae* <SEQ ID 3587> which encodes the amino acid sequence <SEQ ID 3588>. Analysis of this protein sequence reveals the following:

```

      Possible site: 29
      >>> Seems to have no N-terminal signal sequence

20  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1508 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25  The protein has homology with the following sequences in the GENPEPT database.
      >GP:BA06505 GB:AP001516 unconserved protein [Bacillus halodurans]
      Identities = 42/107 (39%), Positives = 70/107 (65%), Gaps = 4/107 (3%)

      Query: 5 NLGEIVISPRVLEVTIGIAANTKVDGVHSLRNK---AVTDSLKKSLGRGVYLNKEEDDT 61
      +LG + ISF V+KVI GIAR++V+GV ++R V + L K+ G+GV ++ D+ +
30  Sbjct: 15 DLGRVEISPEVIEVINGIAASEVEGVATMRGNFAGVARKLGYKNHGKGVKV-DLADDEGI 73

      Query: 62 AADIYVYLQYGVNVPVAVSIAIQQAVKTAVYDMAEVKISSVNIHVGEI 108
      D+ V + YGV+VP V+ IQQ +K A+ M +++ S+H+H+ G+
35  Sbjct: 74 IVDVSVIILYGVSVPEVAKIQQMIKQALQMTAIELQSLNVIHVGV 120

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3589> which encodes the amino acid sequence <SEQ ID 3590>. Analysis of this protein sequence reveals the following:

```

      Possible site: 41
      >>> Seems to have no N-terminal signal sequence

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0882 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 101/129 (78%), Positives = 113/129 (87%)

50  Query: 1 MITEINLGRIVISPRVLEVTIGIAANTKVDGVHSLRNKAVTDSLKKSLGRGVYLNKEEDDT 60
      MTTE +GEIVISPRVLEVTIGIA T+V+GVHSL NK + DS +K SLG+GVYL+ EED +
      Sbjct: 1 MTTEYIGEIVISPRVLEVTIGIATTVQVGVHSLHNKMDSPFNKASLGRGVYLTQEDDGS 60

      Query: 61 VAADIYVYLQYGVNVPVAVSIAIQQAVKTAVYDMAEVKISSVNIHVGEIVPEKPKDELKS 120
      V ADIYVYLQYGV VP VS+ IQ+ VK+AVYDMAEV IS+VNIHVGEIV ETEKPKDELKS
55  Sbjct: 61 VTADIYVYLQYGVKVPVEMNIQKTVKSAVYDMAEVPTISAVNIHVGEIVAEKTPKPKDELKS 120

      Query: 121 LFDEDFLDD 129
      LFDEDFLDD

```

Sbjct: 121 LFDEDFLDD 129

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1158

A DNA sequence (GBSx1234) was identified in *S.agalactiae* <SEQ ID 3591> which encodes the amino acid sequence <SEQ ID 3592>. This protein is predicted to be a utilization substance protein b homolog (nusB). Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.32 Transmembrane 48 - 64 (47 - 64)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14363 GB:Z99116 similar to transcription termination
 [Bacillus subtilis]
 Identities = 51/129 (39%), Positives = 82/129 (63%), Gaps = 9/129 (7%)
 Query: 9 RRDLRERAFQTLFSLETGGEPIDAAHPAYGYDKTVSEDKVLEVPFLNMLNVGVHDHKE 68
 RR RE+A Q LF ++ ++ A + + E+K F LV+GV++H+D+
 25 Sbjct: 3 RRTAREKALQALFQIDVSDIAVNEA-----IEHALDEEKT---DFFEQLVHGVLEHQD 54
 Query: 69 LDTLISSHLKGNSLERLTLDVKSLRLGLYEIKYFDKTPDPRVALNEIIEAKKYSDETS 128
 LD +IS HL + W L+R+ VD+++LRL YE+ Y ++ P V++NE IE+AK++ D+ +
 Sbjct: 55 LDEMISKHLVN-WKLDRIANVDRAILRLAAYEMAYARDIPVNVSMNEATELAKRFGDDKA 113
 30 Query: 129 AKFVNGLLS 137
 KPVNG+LS
 Sbjct: 114 TKFVNGVLS 122

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3593> which encodes the amino acid sequence <SEQ ID 3594>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 53 - 69 (53 - 69)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14363 GB:Z99116 similar to transcription termination
 [Bacillus subtilis]
 Identities = 47/134 (35%), Positives = 76/134 (56%), Gaps = 10/134 (7%)
 Query: 15 RRDLRERAFQALFNIEBMGAELLAASQFAYGYDKTVGEDAQVLELPIFLSLVTVGVNHKE 74
 RR RE+A QALF I++ +++ + D+ + F LV GV H++
 Sbjct: 3 RRTAREKALQALFQIDV-SDIAVNEATEHALDEEKTDP-----DFFEQLVHGVLEHQD 53
 55 Query: 75 ELDNLTSLHLKGNSLERLTLDTKLRLGLYEIKYFDKTPDPRVALNEIIEVVKYSDET 134
 +LD +IS HL W L+R+ D+ +LRL +E+ Y + P V++NE IE+K++ D+
 Sbjct: 54 QDEMISKHLVN-WKLDRIANVDRAILRLAAYEMAYARDIPVNVSMNEATELAKRFGDDKA 112

-1295-

Query: 135 SAKFINGLLSQYVS 148
 + KF+NG+LS S
 Sbjct: 113 ATKFVNGVLSNIKS 126

- 5 An alignment of the GAS and GBS proteins is shown below.

Identities = 104/142 (73%), Positives = 125/142 (87%), Gaps = 1/142 (0%)

Query: 1 MTSVFKDSRRDLREERAFQTLPSLETGGEFIDAAHFAYGYDKTVGRD-KVLEVPFIPLNLV 59
 MT- P++SRDLREERAFQ LF+E G E + A+ FAYGYDK ED +VLE+PIPLL+LV
 10 Sbjct: 7 MTNSPQNSRRDLREERAFQALFNTEGAEALLAAGQFAYGYDKTVTGEDAQVLELPFIPLSLV 66

Query: 60 NGVVDHKEDELDTLISHLKSGWSLERLTLDVKSLLRLGLYETKYFDSTPDORVALNRIET 119
 GV +HK+ELD LIS+HLK GWSLERLT DK+LLRLGL+ETKYFD+TPORVALNRIET+
 15 Sbjct: 67 TGVNNHKEELDKLISHLKSGWSLERLTLDTKTLRLGLPEIKYFPKTPORVALNRIET 126

Query: 120 AKKYSDETSAKFVNGLLSQYIT 141
 KKKYSDTSAKF+NGLLSQ++
 Sbjct: 127 VKKYSDETSAKFINGLLSQYVS 148

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1159

A DNA sequence (GBSx1235) was identified in *S. agalactiae* <SEQ ID 3595> which encodes the amino acid sequence <SEQ ID 3596>. Analysis of this protein sequence reveals the following:

25 Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.81 Transmembrane 239 - 255 (239 - 255)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC31628 GB:U46902 ScrR [Streptococcus mutans]
 Identities = 225/320 (70%), Positives = 273/320 (85%)

Query: 1 MVAKLTDVAALAGVSPITVSRVINKGYLSQKTVTKVNEAMRTLYGKPNILARSLOGKSA 60
 MVAKLTDVA LAGVSPITVSRVIN+KGYLS+KT+TKV AM+TLGYKPNILARSLOGKSA
 40 Sbjct: 1 MVAKLTDVAKLAGVSPITVSRVINRGYLSKRTITTKVQAAMKTLGYKPNILARSLOGKSA 60

Query: 61 KLIGLIFPNIRIFYAEILHLELFLKHYKVTILCNSEKDPKEKEYLEMLGANQVDGI 120
 KLIGLIFPNI +IPY+ELIE+LELFLFKHYK I+CNH+ +P KE++YLEML ANQVDGI
 45 Sbjct: 61 KLIGLIFPNISHIPYSELIEVLELFLFKHYKAITCNHNNPDKERDYLEMLGANQVDGI 120

Query: 121 ISSSHNLGIDDEYKVEAPIAFDNLAPHIPIVSSDNFPGGKMAAQTLLKKHGCGQKMIT 180
 ISSSHNLGIDDEYKVP API+AFDNLAP+IPIVSSDNF GG+MAA+ LKKGCGQ IMI
 50 Sbjct: 121 ISSSHNLGIDDEYKVSAPIAFDNLAPNIPIVSSDNFPGGMAAQLKKHGGCGHPINIA 180

Query: 181 GNDNSDPTGLRRLGFSYSEKSKVITVINGLNNMRREMLKSIISTHKGDIPTSDDLT 240
 G DNS+SPT LR+LGF ++ + ++ LS +R+EMK K I+ KPDGIF SDD+T
 55 Sbjct: 181 GKDNSNSPTALRQLGFKSVFAQAPIFHLGSELSTIRKEMEKIVILGNKSPDGIFLSDMT 240

Query: 241 ALLVILKISQLGLSIPEDIKIVGVDGTSFIQDYVPHLTITIKQPIREIAQLMVEILLAKIE 300
 A+L +K+ +QL ++IP ++K+IGYDGT P++Y P+LTTI+QPI++IA L+V+ILL KI+
 55 Sbjct: 241 AILIMKIANQLNITIPHLKLIQYDGTDFVENYYPYITITIRQPKIDIAHLVDILAKID 300

Query: 301 GQKINKDYILPVSLIGSSV 320
 Q KDYLPLV L+ G SV
 60 Sbjct: 301 HQDIPKDYILPVGLSGBSV 320

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3597> which encodes the amino acid sequence <SEQ ID 3598>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>OP:AAC31628 GB:U46902 ScrR [Streptococcus mutans]
 Identities = 226/321 (70%), Positives = 269/321 (83%), Gaps = 1/321 (0%)
 Query: 1 VVAKLTDVAALAGVSPITVSRVINKKGYSQKTVNKVKNKAMRELGYKPNILARSLQKST 60
 +VAKLTDVA LAGVSPITVSRVIN+KGYS+KT+ KV AM+ LGYKPNILARSLQKGS
 Sbjct: 1 MVAKLTDVAKLAGVSPITVSRVINRKGYSSEKTTIKVQAAKMTLGYKPNILARSLQKSA 60
 Query: 61 QLIGLIFPNISNI FYAELIEHLEIELFKQGYKTTICNSEHNFPKEREYLEMLANQV DGI 120
 +LIGLIFPNIS+IFY+ELIE+LEIELFK GYK IICNS++NP KER+YLEML ANQVDGI
 Sbjct: 61 KLIGLIFPNISHI FYSELIEHLEIELFKHGYSKALICNSQNNPDKERDYLEMLANQV DGI 120
 Query: 121 ISSSHNLGIDYERVEAPIVAFDNLAPNIPVSSDNFEGGKLAQTLKHGGQINVMIT 180
 ISSSHNLGI+DYE+V API+AFDNLAPNIP++SSDNFEGG++AA+ L+KHGGQ+ +MI
 Sbjct: 121 ISSSHNLGIDYERKVSAPITAFDNLAPNIPVSSDNFEGGRMAAKLKHGGQHPIA 180
 Query: 181 GNDNSDSPGTLRQLGPNYQLKRSASIKLENNLSPVRRMEIKSILATRKPDOLFVSDDL 240
 G DNS+SPT LRQLGF + A I L LS +R+EMEIK IL KPDG+P+SDD+
 Sbjct: 181 GKDNSNSPTALRQLGFK-SVFAQAPIFHLSGELSIRKEMEIKVILQNKPDGFI SDDL 239
 Query: 241 TAILIMKVAQLHITIPEDMKVIGYDGTTFIQYVPLATIRQPIDEIAKLSVEILLK 300
 TAIL MK+A QL+ITIP ++K+IGYDGT F++ Y P L TIRQPI +IA L V+L+KKI
 Sbjct: 240 TAILIMKIANQLNITIPHEKLKIGYDGTDFVENYYPVLTITRQPIKDIAHLVDILLK 299
 Query: 301 KKEKTSKDYILPITLLPGASI 321
 + KDYILP+ LL G S+
 Sbjct: 300 DHQDIPKDYILPVGLSGESV 320

An alignment of the GAS and GBS proteins is shown below.

Identities = 247/321 (76%), Positives = 293/321 (90%), Gaps = 1/321 (0%)
 Query: 1 MVAKLTDVAALAGVSPITVSRVINKKGYSQKTVTKVKNRAMELTGYKPNILARSLQKSA 60
 +VAKLTDVAALAGVSPITVSRVINKKGYSQKTV KVN+AMR LGYKPNILARSLQKGS
 Sbjct: 1 MVAKLTDVAALAGVSPITVSRVINKKGYSQKTVNKVKNKAMRELGYKPNILARSLQKST 60
 Query: 61 KLIGLIFPNIRNI FYAELIEHLEIELFKHGYSKTTICNSEKDPKEREYLEMLANQV DGI 120
 +LIGLIFPNI NI FYAELIEHLEIELFK GYKTI+CNSR +P+KE+EYLEML ANQVDGI
 Sbjct: 61 QLIGLIFPNISHI FYSELIEHLEIELFKQGYKTTICNSEHNFPKEREYLEMLANQV DGI 120
 Query: 121 ISSSHNLGIDYERVEAPIVAFDNLAPNIPVSSDNFEGGKMAQTLKHGGQINVMIT 180
 ISSSHNLGI+DYE+VEAPIVAFDNLAP+IP++SSDNF GKG+AAQTL+KHGGQ ++MIT
 Sbjct: 121 ISSSHNLGIDYERVEAPIVAFDNLAPNIPVSSDNFEGGKLAQTLKHGGQINVMIT 180
 Query: 181 GNDNSDSPGTLRRLGFSYSEKES-KVITVTNGLSNMRREMEIKSIISTHKPDGFIPTSD 239
 GNDNSDSPGTLGR+LGF+Y+ K S ++I + N LS +RREME+KSI++T KPDG+P SDDL
 Sbjct: 181 GNDNSDSPGTLRQLGPNYQLKRSASIKLENNLSPVRRMEIKSILATRKPDOLFVSDDL 240
 Query: 240 TALLVIKLSIQLGLSIPEDIKVIYDGTSTFIQYVPHLITIRQPIREIAQLINVEILLAK 299
 Ta+L+K+ QL ++IPED+KVIYDGT+FIQ YVP L TI+QPI BIA+L VEIL+ KI
 Sbjct: 241 TAILIMKVAQLHITIPEDMKVIGYDGTTFIQYVPLATIRQPIDEIAKLSVEILLK 300
 Query: 300 EGQKTKNDYILPVSLINGSSV 320
 + +KT+KDYILP++L+PG+S+
 Sbjct: 301 KKEKTSKDYILPITLLPGASI 321

-1297-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1160

- 5 A DNA sequence (GBSx1236) was identified in *S. agalactiae* <SEQ ID 3599> which encodes the amino acid sequence <SEQ ID 3600>. This protein is predicted to be sucrose-6-phosphate hydrolase (cscA). Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4775 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA35872 GB:XS1507 sucrose-6-phosphate hydrolase [Streptococcus mutans]
Identities = 303/479 (63%), Positives = 359/479 (74%), Gaps = 25/479 (5%)

20 Query: 1 MNLPTETIRYRFYDEWTEHKKENIVKIVSKSPWRATYHLEAKTGILNDPNGFSYFNGKFL 60
MNLPT IRRY Y +WTEE+ ++I NV+ SPW TYH+E KTGILNDPNGFSYFNGKFL+ L
Sbjct: 1 MNLPCNIRYRRYQDWTEEREIKSIKTNVALSPWHTYTHIEPKTGILNDPNGFSYFNGKFL 60

25 Query: 61 FYQNMFFGAAGHLKQWHTESDDLVHFKETGILKLPDVRNDSHGAYSGSALADLKLFLF 120
FYQNMFFGAAGHLK W+HTES+DLVHFKETG L PD +DSHGAYSGGA I D+LFLF
Sbjct: 61 FYQNMFFGAAGHLKSMIHTESDILVHFKETGTVLYPDTSDSHGAYSGSAYBIQDLFLF 120

Query: 121 YTGNVDRMIGWRDPRQIGAWMNDGKITKFDKVLISQPNVDTEHFRDPQIFNYDNQFVAV 180
YTGNVDR W R P QIGA+M G I K F VLI QPNVDTEHFRDPQIFNY QFYA+
30 Sbjct: 121 YTGNVDRDKWVRHPLQIGAFMDKKGNIQKFTDVLKQPNVDTEHFRDPQIFNYKQGFAT 180

Query: 181 IGAQNSKKCGFKILYKALNMDIHHNEFVQDLDPGGTGESEYMIECFNII PVKGGKPVLLYSP 240
+GAQ+ LDPGG+ SEYMIECFN++F+ +PVL+YSP
35 Sbjct: 181 VGAQS-----LDPGGKSEYMIECFNIVFINBQPVLLYSP 215

Query: 241 QGLDKNELDYQNIYFNTYKIGQYFDANGSKIVEPSPIYNLDYGFATATQGFNTSDGRAF 300
QGL K+ELDY NIYFNTYK+ Q FD +V+ S I NLD+GFE YATQ FN DGR +
Sbjct: 216 QGLSKSELDYHNIYFNTYKVCQSPDTEKPAVLDAISEIQLNDPGFECTATQAFNIAQGRVY 275

40 Query: 301 IVSNIGLPLDIDYPSDQFDYOGAMSLVKELSIKGNLYQYFVPMKGNLQHQAEKFTQLQT 360
VSNIGLPLDIDYPSD +DYQGA+SLVKEL+K+G LYQYFV A+++LR + + +T
Sbjct: 276 AVSNIGLPLDIDYPSDSYDQGLSLVKELSLJGKLYQYFVAVESLRSEKAVNTYKPET 335

45 Query: 361 NNITYELELLVPRNMLSSPVLFAFNPQQGLSITIDTVKGVKIIDRSQAQGYATFQTSRQ 420
NNITYELEL ++ +LFA+ KG GL+IT+DT G ++IDRS+AG+QYA EFG+ R
Sbjct: 336 NNITYELELLFDSSSVNELLLFADNKGNGLAITVDTKMTILLDRSKAGQVLEFSGQRS 395

Query: 421 CDIPKDATSINIPIDKSIIFEIPINKGEKVPTGRVFPDABQSGIQLKEGHVHGKIFYELKY 479
C I T +NIP+DKSIFEIPINKGEKVPTGRVFP++ +Q+GI +K G G Y+ELKY
50 Sbjct: 396 CSIQAKETVNNIPVDKSIIFEIPINKGEKVPTGRVFPNDKCTGIVIKSGKPSGNITYELKY 454

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3601> which encodes the amino acid sequence <SEQ ID 3602>. Analysis of this protein sequence reveals the following:

- 55 Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4629 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1298-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/479 (60%), Positives = 367/479 (76%)

```

5      Query: 1  MNLPTETIRYRFYDEWTEEDKKNIVKNSVSWRATYHLSAKTGLLNDENGFSYFNGKPHL 60
      M+LP IRYRYP EW+ +D + I + +++SPW + +H+E KTGLLNDENGFSYFNG+HL
      Sbjct: 2  MDLPQAIRYRPFYKEWSSKDYQAITTKMAQSWESQPHVEPKTGLLNDENGFSYFNGRYHL 61

10     Query: 61  FYQNNPFGAAGLQKQVHTASDOLVHFKEGTIKLKPDRVNDHGAYSSALAIIDDLFLF 120
      FYQNNP+GAAGLQKQVH S DLVHF ET +L PDH +DSHGAYSSGA AIDDLFLF
      Sbjct: 62  FYQNNPFGAAGLQKQVHMTSTDLVHFETRSRLLPDRHDSHGAYSSAYAIIDDLFLF 121

15     Query: 121 YTGNVDRMEONRDERQIGAMMINDGKITKPKVLISQPMDVTEHFRDPQIFNYDNQFYAV 180
      YTGNVDR W R P Q+GAWM G I+K +VLI QP+DVTEHFRDPQ+F+Y QFYA+
      Sbjct: 122 YTGNVDRANVVRTPLQVGAMMDRQGNISKI QVLIQEPDVTEHFRDPQLFSYOGQFYAI 181

20     Query: 181 IGAQNSKIKOGFKILYALANDIHIMEFVGDLDOPGGTSGSYMIECPNIIIFVKGKPVLLYSP 240
      IGAQ G IKLYR++N + +W F+ DLDF +G+ETMIECPN++FV KPVLL++SF
      Sbjct: 182 IGAQGLDGRGKILYAVDANVDMNRFIADLDLDDSGTETMIECPNLFVFDVKVLI FSP 241

25     Query: 241 QGLDNEELDYQNIYRATYTKIGQYFDANSSKIVEFSPIYNDYGFAYATQGFNTSGRAF 300
      QGL K +IAYQNIYRATYTKI + F+ + +++ + NLD+GFAYATQ F++ DGR
      Sbjct: 242 QGLAKADLDYQNIYRATYTKIFSPNPETGLGGALQDLDFGFAYATQAFSPFGCVL 301

30     Query: 301 IVSWIGLPEIDIDYPSDQFDYQGAMSLVKELSIKXNLYQYFVPMKNLQHQAEFKTLQT 360
      VSWIGLPEIDIDYF+D++DYQGA+SLVKEL IK+G LYQ FV A++NLR F ++ +
      Sbjct: 302 AVSWIGLPEIDIDYPTDRYDYGALSIVKELRIKNDGILYQTFVSAQLNLRGPAELFHKIDS 361

35     Query: 361 NNTYRLELLVFRNDLSSVLFANPKGGQSITIDYVKGKVIIDRSQAGQYATFPTSRQ 420
      +N YELEL +P +LFA+ KG GL + +HT KG++ IDRS+AG QTA ++GT R
      Sbjct: 362 SNCYBLELITPGQKILDLLEFADQKNGVLRKLVDTTNGQLSIDRSAGVQYATQYGTIVRS 421

40     Query: 421 CDIPKDATSINIFIDKSIPIFIPNGKGVFTGRVFPDABQSGIQLKGHVHGKYPKLY 479
      C IP+ ++N+++D SI EIPIN+G+KV T RVFP Q+GIQ+ E2 G Y+E+Y
      Sbjct: 422 CQIPQGHVILNYYVDNSILEIFINQGVKLTGRVFPFHGQTGIQVVEGQAFGHYEMRY 480

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1161

A DNA sequence (GBSx1237) was identified in *S. agalactiae* <SEQ ID 3603> which encodes the amino acid sequence <SEQ ID 3604>. Analysis of this protein sequence reveals the following:

```

45     Possible site: 14
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2204 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1299-

Example 1162

A DNA sequence (GBSx1238) was identified in *S. agalactiae* <SEQ ID 3605> which encodes the amino acid sequence <SEQ ID 3606>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -7.64    Transmembrane    259 - 275 ( 250 - 283)
INTEGRAL    Likelihood = -4.41    Transmembrane    113 - 129 ( 109 - 130)
INTEGRAL    Likelihood = -3.03    Transmembrane    180 - 196 ( 180 - 196)
INTEGRAL    Likelihood = -3.03    Transmembrane    439 - 455 ( 438 - 456)
INTEGRAL    Likelihood = -2.81    Transmembrane    298 - 314 ( 298 - 317)
INTEGRAL    Likelihood = -2.02    Transmembrane    396 - 412 ( 395 - 412)

----- Final Results -----
bacterial membrane --- Certainty=0.4057 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA09320 GB:AF059741 sucrose-specific PTS permease [Clostridium
beijerinckii]
Identities = 235/453 (51%), Positives = 312/453 (67%), Gaps = 15/453 (3%)

Query: 7   IAKVQVINAIGGASINVESVAHCATRLRVVMVKDETVIDIQNTVENIEKVQGAFFNSGQVQIIF 66
+AK+++ IGG N++SV HCATRLR+++ D+ I++ +ENI+ V+G PF++ QYQII
Sbjct: 6   VAKETILENIGGKENIKVEHCATRLRLILNDKEKINEKATRNIDGVKQPFSAQYQIIL 65

Query: 67   GTGTVNKIYDEVVAQGLPSTSDSQKAEAAQGNAPAIRFTGDFVFPILLPAIVATGLF 126
GTG VN++YD +V Q      T + K EA Q      Q+ RTFGDVFVP++P +VATGLF
Sbjct: 66   GTGPNVEYDVIYVQNSDLN-TGNKKREAYSQMLTIQIKSRTGDFVFPVPIIPVLVATGLF 124

Query: 127  NGRIGAINNETVLALPGITSKAFSSNFYTYTVLTIDTIAFAFFPALISWSAFRVFGNPV 186
NG-RG+ N V      + NF +T VLTDTAFAL PAL++WS + FGG PV
Sbjct: 125  NGLRGILLNLGVQM-----NENFVLFTQVLITDTIAFALPALVANSTMKFGTGPV 174

Query: 187  IGLVLGLMMNSALPHANAVASGDAHPKF--FGF-I PVVGYQNSVLPAFFVGLLGALE 243
IG-V+GLM+V+ +LPHN+AVA+G A PI      G IPVGYQ SVLEA +G++ AK +
Sbjct: 175  IGIIVGLMLVSPSLPHNAVAAAGTATPINLTILGLNI PVVGYQGSVLPAVLGLIAAKTQ 234

Query: 244  KWLHKKIPDVLDDLVLVPLTFTVMISILALFVIGIPIHSEVMVVLAGTKFVLMLPLGLSGL 303
K L K +FDVL++ PP+T      +L L ++GPI H+ E +      K + LP GL GL
Sbjct: 235  KALKKVVFDVLDTVPITLILFSMVLGILLVGPIMENAEQLFGAIGFMGLPFGLSGL 294

Query: 304  ILGGVHQIIVVTGVHIIINLEAQILAADGKDPFNRIITAAITAQAGATLAVGVKTIQNK 363
++GGVHQ+IVVTGVH N LE +L+++ GKD FNR+IT + AQ A LAV VKTK+KK
Sbjct: 295  VVGGVHQIIVVTGVHIALNALVELLSTGKDAFNAMITGQIVAVQGAALAVAVKTKDKK 354

Query: 364  LKALAPPAALSAGLGITEPAIFGVNLRFGKPFIMGLIAGAGMNLISILKLAGTSGFGITI 423
+L L +A A LGITEPAIFGVNLRFP KPII G      GA GG L+ IL LAGT GIT
Sbjct: 355  KRLYITSSAIPALFGITEPAIFGVNLRFP KPIIFGCGAGVGMLSGILHLAGTGMGIT 414

Query: 424  IPGTILYLNGQIVKYILIMVIGTISLAPVLTMYF 456
+PG LLY+N + Y+++ +A+ LF T F
Sbjct: 415  LPMILLYVN-NLGSYILNVNVAIAVAFCLTLEF 446

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3607> which encodes the amino acid sequence <SEQ ID 3608>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -4.99    Transmembrane    111 - 127 ( 108 - 129)
INTEGRAL    Likelihood = -4.57    Transmembrane    176 - 192 ( 176 - 193)
INTEGRAL    Likelihood = -4.35    Transmembrane    436 - 452 ( 431 - 453)
INTEGRAL    Likelihood = -3.88    Transmembrane    295 - 311 ( 293 - 314)
INTEGRAL    Likelihood = -3.50    Transmembrane    259 - 275 ( 253 - 277)

```

-1300-

INTEGRAL Likelihood = -2.07 Transmembrane 405 - 421 (405 - 421)
 INTEGRAL Likelihood = -0.43 Transmembrane 219 - 235 (219 - 235)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.2996 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- 10 >GP:AAC99320 GB:AF059741 sucrose-specific PTS permease [Clostridium
 beijerinckii]
 Identities = 234/451 (51%), Positives = 312/451 (68%), Gaps = 11/451 (2%)
- 15 Query: 1 MNRQIAAEVIALGGRENVRVAHCATRLRMVYDEBKIDKEKASAIKDVGAFFNSGQ 60
 M + +A E++E +GG+EN++SV ECKTRLR+++ D+ KI+++ E ID VKG FF++ Q
 Sbjct: 1 MKEQIVAKEILENIGGKNIKSVKCATRLRLILNDEKINEKAIENIDVGQGFSSAQ 60
- 20 Query: 61 YQMFQGTGTVNNIYDEVVALGLPTSSSTSEKQAKGKGNIFQRAIRTPGDVFPVPIPAIV 120
 YQ+I CTG VN +YD +V T K EA + Q+ RTGDFVFPVPIIP +V
 Sbjct: 61 YQIILGTGTVNNEVDIVGQNSDLV-TGRNCEAYSCMTLIQIKSTPGDVFVPIIPVLV 115
- 25 Query: 121 ATGLFMGVRGLVATQPAIMDLGVHVEYGENFLMYTRILITDTAFVYLPALVANSAPRVFGN 180
 ATGLFMG+RGL+T + + ENF++TV+LTDTF +LPALVMS + PG
 Sbjct: 120 ATGLFMGLRGLLTMVG-----QNNENFVLFYQVLTDTAFAPLPAVMSMKFGGT 172
- 30 Query: 181 PIIGIVLGLMVLNSELPMNVAWVAGSGDK-PLTFPGP-VPVVGYYQTVLPAPFVGLGAK 238
 P+IGIV-GLMDVS LPNA+ VA+G LT G +PVVGYQG+VLPA+G+ AK
 Sbjct: 173 PVIGIVGLMDVSPSLPNAYAVAAGTATPIMLTILGLNIPVVGYYQSVLPALVIGLIAAK 232
- 35 Query: 239 LEKWLHKKVPKPEALDLVTPPLTFAINSTGLFVIGVPHSLNVLVAGTQAVLHFGIA 298
 +K L K VP+ LDL+VTFP+T LGL ++GP+ H+ E L+ + + LFG+
 Sbjct: 233 TQKALKKVVVPALDLIVTPPTIFLESVILGLLVGPIHNAQLIPGAIKGMGLDPGLG 292
- 40 Query: 299 GLVVGIGQQLIVVTGIIHIFNLEAQLANTGDKPFNAYLTAAACAGATLAVAVTKTS 358
 GL+VGG+ QLIVVTG+HH N LE +L+++YKGD FNA +T AQ A LAIVAVTK
 Sbjct: 293 GLVVGGVHQLIVTVGVHNALEVELLSTGKDAFNAMITGSIVAQGAALAVAVTKTD 352
- 45 Query: 359 TKLKLAPFSTLSALGITPAIGVNLRYPKVFSGLIGLGGWAGLFGIAGTGGI 418
 K + L S + A LGITPAIGVNLRY K P+ G GSA+G+G+ +AGTG GI
 Sbjct: 353 KKKRSLYISSAIPFLGITPAIGVNLRYPIKFPFPGCAGVGMGLSGLHLAQTGGI 412
- Query: 419 TVLPGTLLYINQLLQYIVIMLVGLGVAF 449
 T LRG LLY+N L Y++ +V + VAF +
 Sbjct: 413 TALPGMLLYVN-MLGSYILVNVVAIVAFCL 442

An alignment of the GAS and GBS proteins is shown below.

- Identities = 409/618 (66%), Positives = 491/618 (79%), Gaps = 12/618 (1%)
- 50 Query: 4 NTEIAKQVINAIGGASNRVVAHCATRLRMVYDEBKIDKEKASAIKDVGAFFNSGQ 63
 N+IA +VI A+GG NVRSVAHCATRLRMV DE IDK E I+KV-GAFFNSGQ
 Sbjct: 3 NRQIAAEVIALGGRENVRVAHCATRLRMVYDEBKIDKEKASAIKDVGAFFNSGQ 62
- 55 Query: 64 IIRGTGTVNKIYDEVVAQGLPTSSSTSDQKASAAQNSAPQRAIRTPGDVFPVPIPAIVAT 123
 +IRGTGTVN IYDEVVA GLPTSSST+QKAA K GN FQRAIRTPGDVFPV+PAIVAT
 Sbjct: 63 MIFGTGTVNNIYDEVVALGLPTSSSTSEKQAKGKGNIFQRAIRTPGDVFPVPIPAIVAT 122
- 60 Query: 124 GLFMGIRGINNDTVIALFGTTTSKAPSSNFYITVVLITDTAFAPVLPALVANSAPRVFG 183
 GLFMG+RG + ++ LFG NF YT +LTDTF + PAL+MSAPRVFG
 Sbjct: 123 GLFMGVRGLVATQPAIMDLGVHVEY-----ENFIMYTRILITDTAFVYLPALVANSAPRVFG 179
- 65 Query: 184 NPVIGLVLGLMVLNSELPMNVAWVAGSGDK-DHPIFGFGFIPVVGYYQNSVLPAFPVGLGAKL 242
 NP+IG+VLGLM+V++ LPNMA VASG D P+ FFGF+PVVGYQ +VLPAFPVGL+GAKL
 Sbjct: 180 NPIIGIVLGLMVLNSELPMNVAWVAGSGDKPLTFPGFVPVVGYYQTVLPAPFVGLGAKL 239
- Query: 243 EKNWLHKKIPDVLDDLIVFFLITFTVMSIILAFVIGPIFHSVENVLAGTKFVNLPLGLSG 302
 EKNWLHKK+P+ LDLL+ FFLTF +MS L LFVIGP+PHS+EN VLAGT+ VL+LIP G++G

-1301-

Sbjct: 240 EKWLEHKVPEALDILLVTPFLTFAINSTLGLFVIGVPVHSENLVLAQTQAVLHLPFGIAG 299

Query: 303 LILGSGVHIQIVVTGVHHIFNLEAQLIADGKDPFNIAITAMTAQAGATLAVGVKTEK 362
LI+GG+G+IVVTG+HHIFN LEAQLIA GKDPFNA +TAA AQAATLAV VKTE+

5 Sbjct: 300 LIVGSHIQLIVVTGIHHIFNLEAQLIANTGKDPFNAYLTAAATAQAGATLAVAVKTEK 359

Query: 363 KKLALFAPALASAGLGITEPAIFGVNLRGKPFIMGLIAGAAGGWLASILKLAGTGFGIT 422
KLK LAPP+ LSA LGITEPAIFGVNLR+ K F+ GLI GA GSW+A + +AGTGFGIT

10 Sbjct: 360 KKLGLAFSTLSALILGITEPAIFGVNLRYPKVPVSLIGQALGQWAGLFGIAGTGFGIT 419

Query: 423 LIPTGLLYNLNGQIVKYLIMVIGITSLAFVLITMFGYEDKDKSAVARVSPVSEETLDDPTI 482
++ZGTLLYNLNGQ++YL+ ++ +AF +Y +GY+D++ + V V++T D P +

15 Sbjct: 420 VLEPTLLYNLNGQLLYLVIMVLGLGVAFATYTWGYQDRETLLEAVS--VDQTADQPAL 477

Query: 483 TQTSQLPARTIVSPIDGQVIALDVTSDPVFSSGIMDGLAIKPRGNTIYSPVDGFVQIAF 542
+ ET+ SFL+G V+ L VSDPVFSSG MG GLAIKP NT+YSPVDG V+I F

20 Sbjct: 478 AE-----ETLTSPLNGITVVDLSAVSDPVFSSGAMQGLAIKPEENTLYSPVDGKVEIVF 531

Query: 543 ETGHAYGKSDKGAEILIHIGIDITVNGTGFTSKVKADQKVKKGIDILGTFDSAKIARAG 602
ETGHAY I S +GAE+L+HIGIDIT +M G GF S V Q VKKG+LG FD +KIAEAG

25 Sbjct: 532 ETGHAYAITSSQGAFLVLIHIGIDITSEMAAGDGFPSLVAVGQAVKKGIDILGTFDPSKIAARAG 591

Query: 603 LENTAMIIVTNFADFADV 620
LD+T M+IV+N AD+ V

30 Sbjct: 592 LDDTTMIVSNADYQSV 609

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1163

30 A DNA sequence (GBSx1239) was identified in *S. agalactiae* <SEQ ID 3609> which encodes the amino acid sequence <SEQ ID 3610>. This protein is predicted to be fructokinase. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2436 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA02467 GB:D13175 fructokinase (Streptococcus mutans)
Identities = 232/291 (79%), Positives = 257/291 (87%)

45 Query: 1 MTKLYGSIIAGGTFKVCVAGDRELKVVEMQFPPTTTQSTIKTKTDFKFRPEKKLEAVAI 60
M+KLKYSIIAGGTFKVCVAGDE ++EK+QFPPTTP ETI+KTV FFK+FE L +VAI

Sbjct: 1 MSKLYGSIIAGGTFKVCVAGDENFQLERKQFPPTTTPYSTIEKTFVAFKCFEADLSVAI 60

Query: 61 GSPGPIIDDKSKTYGYITTTPKLHWANVDLGLISKDFNVPPFYPTTDVNSSAYGEVIAR 120
GSPGPIIDID+ S TYGYIT+TPK +WNVVD +GLISKDF +PFYPTTDVNSSAYGE IAR

50 Sbjct: 61 GSPGPIIDIDQNSDTYGYITSTPKPNWNVVDVGLISKDFKIDPIFYPTTDVNSSAYGETIAR 120

Query: 121 NNIDSLVYTYTIGTGICAGAIQKGFIQCTGTRAGHTYMMHMQDQANDPKGICPFHNSC 180
+N+ SLVYTYTIGTGICAGAIQ GEFIG GTRAGH YMA HP D + F G CFFH C

55 Sbjct: 121 SNVKSIVYTYTIGTGICAGAIQNGRFGIGMGHTRAGHYMAHENDVHKGFPVGTCPFHKG 180

Query: 181 LEGLASGPTLEARTGIRGELIENSMVMDVQAYYIAQAIAQATVLYRQVIVFEGGGVMAQ 240
LEGLA+GP+LEARTGIRGELIEN+ S VMD+QAYYIAQAIAQATVLYRQVIVFEGGGVMAQ

60 Sbjct: 181 LEGLASGPTLEARTGIRGELIENQSEVMDIQAAYIAQAIAQATVLYRQVIVFEGGGVMAQ 240

Query: 241 EEMLRKRVQTFATLLNGYLPVPLDSYIVTATENSGSATLGNLAFKKS 291

-1302-

EHML RVR+ F +LLM YLPVVD+ DYIVTPA+ +NGSATLGN ALAKKI+
 Sbjct: 241 EHMLRVRKFKPTSLNDYLPVVDKDYIVTPA+NGSATLGNLALAKKIA 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3611> which encodes the amino acid sequence <SEQ ID 3612>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2012 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 212/293 (72%), Positives = 246/293 (83%)
 Query: 1 NTKLYGSIEAGGTFKVCAGVDEELKVVEMQPFITTPQSTIKKTVDFKPEKKLEAVAI 60
 M KLYGSIEAGGTFKVCAGVDEE VV-K QFTTTP+ETI +T+ +FK FE L +AI
 Sbjct: 1 NGKLYGSIEAGGTFKVCAGVDEEFTVVDKTFQFTTTPETIARTIAYFAEADLAGMAI 60
 Query: 61 GSGPGIDIDKSGKTYGYITTTPEKLHWANVDLLGLSKDHFVPPYFTFDVNSGAYGEVIAR 120
 GSGPGIDID S+TYGYITTTPEK WANVDLLG +S F +PF TTDVNSGAYGEVIAR
 Sbjct: 61 GSGPGIDIDPSSETTYGYITTTPEKSGWANVDLLGLSAAFKIPFDVTTDVNSGAYGEVIAR 120
 Query: 121 NNIDELVYYTICTGIGAGAIQKGEFIGTGRTGRTGHTYMAHPPDQANDEKGCIPFNSC 180
 +SLVYYTICTGIGAGAIQ G FIGG GHTGHTY+ HP D A F G+CPHF C
 Sbjct: 121 PGVBELVYYTICTGIGAGAIQGHGFIGGLGHTGHTYVYMHPPDDMAKGFVGVCPFHGSC 180
 Query: 181 LELGASGPTLEARTGIRGELIERNSNVNDVQAYIYAQAATQATVLYRRQVIVFGGVMQAQ 240
 LELG+A+GP+EARFG+RGE +++ + VND+Q+YIAQA+QAT+LYRRQVIVFGGVMQAQ
 Sbjct: 181 LELGAAGPSIEARTGVRGERLQCEADVNDIQAPYIAQAALQATMLYRQVIVFGGVMQAQ 240
 Query: 241 EHMLRVRKFTATLNGVLPVVDLSDYIVTPA+NGSATLGNALAKKISG 293
 EEM+ RV F LL+GYLPVVDL+DYIVTPA+ +NGSATLGNALAK +G
 Sbjct: 241 EHMVLVRHDKFTALLGGLVLPVVDLDYIVTPA+NGSATLGNALAKLAAGS 293

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1164

A DNA sequence (GBSx1240) was identified in *S.agalactiae* <SEQ ID 3613> which encodes the amino acid sequence <SEQ ID 3614>. This protein is predicted to be Mannosephosphate Isomerase (pmi). Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4717 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA04021 GB:D16594 Mannosephosphate Isomerase [Streptococcus mutans]
 Identities = 232/312 (74%), Positives = 262/312 (83%)

Query: 1 MSEPLFLASMHDKIWGTYKLRDEPGYDIPSETTGGEYMAISAHNPGVSRVINGRFGGCF 60
 M PLFL+ MH KIWG +LR EPGYDIPSETTGGEYMAISAHNPGVS VKNG +KG L
 Sbjct: 1 MEQPLFLQSMHKKIWGNLRKRFGYDIPSETTGGEYMAISAHNPGVSVVKNVYKGVPL 60

-1303-

Query: 61 DKLYQGEKSLFGNPDVTVPFLTKILDANDNLVSVQVHPDDAYALKHEGELGKTECWYIS 120
 D+LY + LFGN + VFPLLTKILDANDNLVSVQVHPD+AYAL+HEGELGKTECWY+IS
 Sbjct: 61 DELYAEHRELFGNPKSSVFPLLTKILDANDNLVSVQVHPDNYALEHEGELGKTECWYIS 120

5 Query: 121 ADEGSEIITYGHNAKTEKRELQMIESGDWEHLLTRIPVKSQDFYVPSGIMHAIGKIGILL 180
 ADEG+EIYGH AK+KEELQMI +GDW+HLTL+IPVK+GDF+YVPSGIMHAIG+GI+IL
 Sbjct: 121 ADEGAETIYGHEAKSEKELQMIAGDWDHLLTKICPVKAGDFYVPSGIMHAIGRSIMIL 180

10 Query: 181 ETQGSDDTTYRVYDFDRPDASGLRDLHIEQSIDVLTIQKANTVPANMLKHLSTLLV 240
 ETQGSDDTTYRVYDFDR D G+R LHIEQSIDVLTIQKAPN PA + L+L +T+LV
 Sbjct: 181 ETQGSDDTTYRVYDFDRKDDQGRKRAIHIEQSIDVLTIQKAPANTPAMLGLQETFTVLV 240

Query: 241 SNDPFTVYKNEISGVTFKQAPYLLVSVLDGAGHITVDNKVYTLKKGDFHFLPNVVKW 300
 S+ FFTVYK+ISG +Q APYLLVSVL G G ITV + Y L+KGDH ILFN + W
 Sbjct: 241 SSPFTVYKNGVSSVVMQQTAPYLLVSVLAGQREITVGLQYALRKGDHILEFTIKSW 300

15 Query: 301 DIDQLEIASH 312
 DG LEIASH
 Sbjct: 301 QPDQLEIASH 312

20 Query: 301 QPDQLEIASH 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3615> which encodes the amino acid sequence <SEQ ID 3616>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3714 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/312 (74%), Positives = 264/312 (84%)

35 Query: 1 MSELPLFLASMHDKINGGTLRDEFGYDIPSETTGEYMAISAHPNGSVRVKGRFGKQPL 60
 MSELPLFL+++MHD+INGGTLKRD F Y+IPG+TTGEYMAISAHPNGVS V NGR++G L
 Sbjct: 1 MSELPLFKSTMHDIRINGGTLKRDVPAYNI PSDTTGEYMAISAHPNGSVSTVTKRGYQQPL 60

40 Query: 61 DKLYQGEKSLFGNPDVTVPFLTKILDANDNLVSVQVHPDDAYALKHEGELGKTECWYIS 120
 + LY E +LFGN + VFPLLTKILDANDNLVSVQVHPDDAY +HEGELGKTECWYIS
 Sbjct: 61 NTLVAQEPALFGNPKEEVFPLLTKILDANDNLVSVQVHPDDAYGREHEGELGKTECWYIS 120

45 Query: 121 ADEGSEIITYGHNAKTEKRELQMIESGDWEHLLTRIPVKSQDFYVPSGIMHAIGKIGILL 180
 A+EGSEI+YGH AK+KE+LR MIE+G W+ LLTR+PVK+GDF+YVPSGIMHAIGKIGILL
 Sbjct: 121 AREGSEIVYGHCAKSEKELRAMIENGAMDLLTRVPVKGDFYVPSGIMHAIGKIGILL 180

Query: 181 ETQGSDDTTYRVYDFDRPDASGLRDLHIEQSIDVLTIQKANTVPANMLKHLSTLLV 240
 ETQGSDDTTYRVYDFDR D +G LRDLEH+SIDVLTIQK N+VPA M L ++ +T LV
 Sbjct: 181 ETQGSDDTTYRVYDFDRKDDVKNLRLHIEKSIDVLTIQKPNVPAIMVLNNVATTLV 240

50 Query: 241 SNDPFTVYKNEISGVTFKQAPYLLVSVLDGAGHITVDNKVYTLKKGDFHFLPNVVKW 300
 S FFTVYK S + + KQ APYLLVSVL G G + VD K Y L+KG HFILPNVD W
 Sbjct: 241 STPFPTVYKMTVSQVDMKQAPYLLVSVLKGQKLYVDQKAYLEKEMFILPNVDKSW 300

55 Query: 301 DIDQLEIASH 312
 DQLE+I SH
 Sbjct: 301 SFDQLEIASH 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1165

A DNA sequence (GBSx1241) was identified in *S. agalactiae* <SEQ ID 3617> which encodes the amino acid sequence <SEQ ID 3618>. This protein is predicted to be preprotein translocase secA subunit (secA).

Analysis of this protein sequence reveals the following:

```

5  Possible site: 53
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.1102 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10107> which encodes amino acid sequence <SEQ ID 10108> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA50286 GB:L32090 secA [Listeria monocytogenes]
Identities = 503/843 (59%), Positives = 643/843 (75%), Gaps = 16/843 (1%)

```

```

Query: 11  MANILRTVIENDKGELKKLKKIADKVDSDYADHMAALSDDEALQAKTPEFKYQNGETLDQ 70
20  MA +L+ + E+ K ++K L++ A ++ + AD AALSD+AL+ KT EFKR Q GETLD
   Sbjct: 1  MAGLLKKIFESGKKDKVTLERKADEI IALADETAALSDDALREKTVFKERVQKGETLDD 60

Query: 71  LLPEAFVAVREASKRVGLGYPYHVGIMG3VLHHGDIENFMRTGEGKTLTATMFPVYLNAIS 130
25  LL EAFV RE +R LGLYP+ VQ+M93VLH +I EN+TGEKTLTAT+PVLN+A+S
   Sbjct: 61  LLVEAFVAREGAKRALGLYPPKVLGM93VLHEDINIAEMKTGEGKTLTATLPVYLNALIS 120

Query: 131  GLGVHVTIVNEYLSTRDATEHGEVISMGLGSLVGINLAASKPFPEKRAYNCDITISTNAEV 190
30  G GVHV+IVNEYL+ RDA EM3 +Y++LGLSVG+NL A S EKREAY CDTITSTN+
   Sbjct: 121  GEGVHVTIVNEYLHREDAEMGVILNIFLGLSVGLANALSTEKREAYACDTITSTNNEL 180

Query: 191  GFDYLDRNMVVRQEMVQRFLNAYALVDEVDVSLIDEARTPLIVSGFVSSEMMQLATPADM 250
35  GFDYLDRNMVAV +E+MVQREL +A++DEVDG+L+DEARTPLI+SG ++ LY RA+
   Sbjct: 181  GFDYLDRNMVVIKEHNVQRFLAFVAVIDEVDVSLIDEARTPLIISGE-AEKSTILVIRANT 239

Query: 251  FVKTLL-NSDDYLIDVPTKTLGSLDGTGIDKAENYFHLNLYDENVALTHYIDNALRNYI 309
40  FV+TL +DY +D+ TK++ L++ G+ E ENYF +NL+DEN +H+I ALA+NY
   Sbjct: 240  FVRTLIEEDYTVDIKTKSVQLTDEGNTGENYFQVENLFDLENTVILHHIAQALKANYT 299

Query: 310  MLNINIDYVVSKEQRIILVDQPTGRITMGRFRPSGLHQAPEAKESVPIQESKTSASITYQ 369
45  M L++DYVV ++ E+LIVDQPTGR M+GRFRPS+GLHQA+RAKE V IQ ESKT A+IT+Q
   Sbjct: 300  MSLDLYDYVV-QDDEVLVDQPTGRIMKGRFRPSBGLHQAERKGVTTIQNSKTMATITPQ 358

Query: 370  NMFMYKLAGMTGTCTKTESEPRDIYNMVRVPIPTNRPVQRIIDHSDLLYPTLDSKFRAY 429
50  N FMRY KLAGMTGT KTESEPR+IYNMVRV IPTN+ + R D DL+Y I+++KF AV
   Sbjct: 359  NYFMRYKLAGMTGTCTKTESEPRDIYNMVRVPIPTNKVIRDRDPDIYTTMEAKFNAY 418

Query: 430  VADVGRYEQCGQFVLKGTAVETSDLSIKLVAAGVPHVHLNAXNHPKEAQLIHAEQGR 489
55  V D+ ER+ +GQFVLKGTAV+ +LIS KL G+ H+VLNAX H+EA II +AG+RG
   Sbjct: 419  VEDIAERHAKQGFVLKGTAVMNI-ELISSKLKRKGIKHVDINAKQHERADIIHAGERG 477

Query: 490  AVVIATNVAAGRGTDIKLGGVRLGGLGVICTERHESRRIDNQLGRGRGQDDPGESQFY 549
60  AV IATNVAAGRGTDIKLGG E GGL VICTERHESRRIDNQLGRGRGQDDPG+QFY
   Sbjct: 478  AVVIATNVAAGRGTDIKLGGTTHAGGLAVICTERHESRRIDNQLGRGRGQDDPGVTCQFY 537

Query: 550  LSLRDDLMMRFRTGTRIKVLERMNLAEDDTVIKSMKLTQVESACRRVVEGNHYDTKQVIL 609
65  Ls+ED+LMRFRG+D +K ++R +AED I+SMM++R VESACRRVVEGNH+D+KQVIL
   Sbjct: 538  LSMEDLMRRFRSSDMMKMMERPMQAE--AIQSKMVSRAVESACRRVVEGNHPSRKQVIL 595

Query: 610  QYDDVMKQREIIYANREVVITAEKDLGPELGMKIKKTKRAVDANSHSDEKFTAA---EA 666
70  QYDDV+R+QRE+IY R EVI AE L ++ MI+RT+ V +++ S + A +
   Sbjct: 596  QYDDVLRQOREVITKQRYEVINAENSLRELTETMCIQRTVNFIVSNASSHEPEZAWNQG 655

```

-1305-

Query: 667 I V N F A R S A L L D E B A I T V S K L G K L K B A R T K E L L Y E R A L A V V I S Q I A K L K D P E A I I F P Q K V L 726
 I++ + L L E T T+ +L+ +I+ L+ ++ A Y++ L P E P+K V+
 Sbjct: 656 I I D Y V D A N I L P E G T I T L E D L Q N R T S E D I Q N L I D K I K A A Y D E K - E T L L P P E F P N E P A V V 714

5 Query: 727 L I M V V N Q W T H E I D A L D Q L A N S V G L R G Y A Q N N P I V E Y Q S G E G F M Q D M I G S I E P D V T R L 786
 +L V V D +W +H I D A +D L R+ + L R Y Q +P+ E Y Q S G F M E+ M+ S I+ D W R +
 Sbjct: 715 L L R V V T K W V D H I D M O H L D G I H L R A Y Q I D P L R E Y Q S G E F M P E A M V S I D E D V A R Y I 774

10 Query: 787 M K A Q I H E Q - E R E R - A S Q H A T T A R Q N I S A Q H V M N N I S P H Y Q I K R N D K C P C G S G K K F K M 844
 M K A+ I + E R E+ A A A R A+ P+ + Q I R N D C P C G S G K K N
 Sbjct: 775 M K A E I R Q N L E R B Q V A K G E A I N P A R G K P E A K R Q P I R K D - - - - Q H I G R N D P C P C G S G K K Y K N 830

Query: 845 C H G 847
 C H G
 Sbjct: 831 C H G 833

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3619> which encodes the amino acid sequence <SEQ ID 3620>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4443 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 710/837 (84%), Positives = 777/837 (92%), Gaps = 3/837 (0%)

30 Query: 11 M A N I L R F V I E N D K G E L K L K I A K K V D S Y A D M A A L S D E A L Q A K T P F K E R Y Q K G E T L D Q 70
 M A N I L R V I E N D K G E L + K L + K I A K V + S Y A D M A + L S D L Q K T F P K E R Y Q G E I L + Q
 Sbjct: 1 M A N I L R K V I E N D K G E L R K L E K I A K K V E S Y A D C M A L S D S R D L Q G K T L E F K E R Y Q K G E T L E Q 60

35 Query: 71 L L P E A F A V R E A K R V L G L P Y P H V Q I M G G I V L H H G D I P E M R T G E G K T L A T M P V Y L N A I S 130
 L L P E A F A V R E A + K R V L G L + P Y V O I M G G I V L H + G D + P E M R T G E G K T L A T M P V Y L N A I +
 Sbjct: 61 L L P E A F A V R E A A K R V L G L P P Y P V Q I M G G I V L H N G D V P E M R T G E G K T L A T M P V Y L N A I A 120

Query: 131 G L G V H V I T V N E Y L S T R D A T E M G E V Y S M L G L S V G I N L A A K S P F E K R E A Y N C D I T Y S T N A S V 190
 G G V H V I T V N E Y L S T R D A T E M G E V Y S M L G L S V G I N L A A K S P E K R E A Y N C D I T Y S T N + E V
 40 Sbjct: 121 G E G V H V I T V N E Y L S T R D A T E M G E V Y S M L G L S V G I N L A A K S P A K R E A Y N C D I T Y S T N S E V 180

Query: 191 G F D Y L R N M V R Q E D M V Q R P L N A L V D E V D S V L I D E A R T P L I V S G A V S S E T N Q L Y T R A D M 250
 G F D Y L R N M V R Q E D M V Q R P L N + A L V D E V D S V L I D E A R T P L I V S G V S S E N Q L Y T R A D M
 45 Sbjct: 181 G F D Y L R N M V R Q E D M V Q R P L N F A L V D E V D S V L I D E A R T P L I V S G A V S S E T N Q L Y T R A D M 240

Query: 251 F V K T L S D D V I D V P K T K I G L S D G I D K A S Y P N L N L Y D I E N V A L T H Y I D A L N A R N I M 310
 F V K T L S D D V I D V P K T K I G L S D + G I D K A + Y P + L + N L Y D + E N V A L T R + I D A L N A R N I M
 Sbjct: 241 F V K T L S D V D V I D V P K T K I G L S D S G I D K A S Y P N L N L Y D I E N V A L T H P I D A L N A R N I M 300

50 Query: 311 L L I D Y V V S E G E I L I V D Q T G R T M G R R F S D G L H Q A I E A K S V P I Q E E S K T S A S I T Y Q N 370
 L L + I D Y V S E R + E I L I V D Q T G R T M G R R F S D G L H Q A I E A K E V I Q E E S K T S A S I T Y Q N
 Sbjct: 301 L L D I D Y V V S E G E I L I V D Q T G R T M G R R F S D G L H Q A I E A K G V R I Q E E S K T S A S I T Y Q N 360

55 Query: 371 M F R M Y K L A G M T G T K T E E E F R E Y V N M R I I P I P T N R P V Q R I D H S D L L Y P T L S K P R A V V 430
 M F R M Y K L A G M T G T K T E E E F R E + Y N M R + I P I P T N R P + R I D H + D L L Y P T L + S K P R A V V
 Sbjct: 361 M F R M Y K L A G M T G T A K T E E E F R E Y V N M R I I P I P T N R P I A K I D H T O L L Y P T L S K P R A V V 420

Query: 431 A D V K E R Y E Q G P V L V G T V A V E T S D L I S R K L V A A G V P H E V L N A K N H F K R A Q I M N A G Q R G A 490
 D V K R + + G Q P + L V G T V A V E T S D L I S R K L V A A G V P H E V L N A K N H F K R A Q I M N A G Q R G A
 60 Sbjct: 421 B D V K T R H A G Q P I L V G T V A V E T S D L I S R K L V E A G I P H E V L N A K N H F K R A Q I M N A G Q R G A 480

Query: 491 V T I A T N M A G R G T D I K L G B V R E L G G L C V I G T E R H E S R R I D N Q L R G S G R G Q D P G S Q P Y L 550
 V T I A T N M A G R G T D I K L G B V R E L G G L C V I G T E R H E S R R I D N Q L R G S G R G Q D P G S Q P Y L
 Sbjct: 481 V T I A T N M A G R G T D I K L G B V R E L G G L C V I G T E R H E S R R I D N Q L R G S G R G Q D P G S Q P Y L 540

-1306-

Query: 551 SLEDDIARRPGTIRIKVFLERMLARDDTVIKSGMLTRQVESAQRRVBCNNYDTRKQVLQ 610
 SLEDDIARRPG+DRIK L+RM L E+D+VIKS ML RQVESAQ+RVBCNNYDTRKQVLQ
 Sbjct: 541 SLEDDIARRPGSDRIKAVFLERMLARDDTVIKSGMLGRQVESAQRRVBCNNYDTRKQVLQ 600

5 Query: 611 YDDVMBQRREIYANRRSVITAKRLGPELKSMIKRTIKRAVDHARSDDKTAASAVINF 670
 YDDVMBQRREIYANRR+VITA RULGPE+K MIKRTI RAVDAH+RS++ A +LIV F
 Sbjct: 601 YDDVMBQRREIYANRRDVTANRDLGPEITAMIKRTIDRAVDHARSNNKDAIDAIVTF 660

10 Query: 671 ARSALLDEEATTVSELRLGLKEATIKRLLYERALAVYRQOIAKLKDPFAITIEPQVLILIMV 730
 AR++L+ EE I+ ELRLGL+ +IKR LV+RALA+Y+QQ++KL+D EAITEPQVLILIM+
 Sbjct: 661 ARTSLVPEEPISAKELRLGLKDDQIKELKYQALAIYDQQLSKLRDQEAITIEPQVLILIM 720

Query: 731 VDNQWTEHIDALDQLRNAVGLRGYAQNNPVVEYQSGSFRMFQDMIGSIEFDVTTMLAQ 790
 VDN+VTEHIDALDQLRNA+VGLRGYAQNNP+VEYQ+EGF+MFQDMIG+IEFDVTT+MLAQ
 Sbjct: 721 VDNQWTEHIDALDQLRNAVGLRGYAQNNPVVEYQAGSFRMFQDMIGSIEFDVTTMLAQ 780

15 Query: 791 IHEQERERASQHATTAEQNIASHQHVPMNNSPEFYQIKRNDKCPGSGGMFKNCHG 847
 IHEQERERASQ ATT A QNI +Q ++ P+ ++RN+ CPCGGG KFKNCHG
 Sbjct: 781 IHEQERERASQRATTAPQNIQSQCANTDLEPK---VERNEACPGSGGKFKNCHG 834

20

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1166

A DNA sequence (GBSx1242) was identified in *S. agalactiae* <SEQ ID 3621> which encodes the amino acid sequence <SEQ ID 3622>. This protein is predicted to be phospho-2-dehydro-3-deoxyheptonate aldolase (aroH). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3429 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF40753 GBS:AE002387 phospho-2-dehydro-3-deoxyheptonate
 aldolase, phe-sensitive [Neisseria meningitidis MC58]
 Identities = 122/348 (35%), Positives = 187/348 (53%), Gaps = 32/348 (9%)

40 Query: 1 MGPHQLSDKINISILKOKTSLDLEVSQKLAKE-----EELKNIIKGEDQRLVIV 51
 M H +D I I + +K+ + +KE +E+ +++ G D+R LVI+
 Sbjct: 1 MTHHYPTDDIKIKEVKELLPTLAIYELFISKEASGLVHRTQETSDLVHGRDRLRLVII 60

45 Query: 52 GPCSAINDPKAVLITYAKRLAKLEAFEDKMPLVMRVYTAKPRTNKGYKGLVHSDKILGVF 111
 GPCS +FKA L YA+RL KL +++++ +VMRVY KPRT G+KGL++ G F
 Sbjct: 61 GPCSHDFKAALYAEKLLKLRKYENELLIVMRVYFEKPRIT-VGWRGLINDPHLDGTF 119

Query: 112 -----FQARKMHYDIIRETGLTDADELLYPEMLSVMDLVSYYAIGARSVEDQSHRPI 165
 QRR + + G+ + E L DL+5+ AIGAR+ E Q HR ++
 50 Sbjct: 120 DINFGLRQARSILLLS-LNNQMPASTEFLEDMITPQYYADLISWAGIARTTESQVHRSL 178

Query: 166 SGIDAFVGNENPTSGNLVMPNNAVYAQNQQLFVYQNKQ-----VRTDGNLSSHVLLEGY 220
 SG+ PVG KN T GNL++ ++A AA + K V T GN HVILRG
 Sbjct: 179 SGLSCFVGNENPTDGNLKIADAIGAASHSHHFLSVTKAGHSATVITGNDPCHVILLEG 238

55 Query: 221 HNADYSRIFPNHYENLLETITHYRETDLPNPFIVDTINDHNSGKQFLEQIRIVKSVLAAR 280
 PNY E++ E + + ++D +H NS K + Q + ++ A
 Sbjct: 239 KE-----PNYDAEVSAAEQLEAGVYTDK-LMIDCSHANSKDYTRMEVAQDIAQL 291

60 Query: 281 QNHTKIRNYVRGFLIESYLEDRQDKPDVFKSITDPCLGMDKTEMLI 328
 + + + G ++ES+L +GRQDKP+V+GKSITD C+WG TE L+

-1307-

Sbjct: 292 E---QDQGNIMGVVVRSHLVBRQDKPEVYKGSITDACTIONGATEELL 336

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3623> which encodes the amino acid sequence <SEQ ID 3624>. Analysis of this protein sequence reveals the following:

5 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1171 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 52/233 (22%), Positives = 93/233 (39%), Gaps = 40/233 (17%)

15 Query: 50 IVGPCCADNPKAVLTAKRLAKLEAAFKDKMFLVMRVYTTAKPHNGDCYKGLVHHSOKLG 109
 IVGPCCS ++ + A KL + R KPRT+ ++GL
 Sbjct: 19 IVGPCCSIRSVDHIRLAASSAKKLGYN-----FRGGAYKPTCSAASFGQLG----- 64

20 Query: 110 VFPQARKMHYDIIRKTEGLTADPELLYFEMLSVMDLVSYTAIGARVEDQGRHFISGID 169
 Q + +++ +E GL+ E++ L D + +GAR+++ S ID
 Sbjct: 65 --LQIRKYLHEVCQSFGLLSVSEIMSERQLEEDYLDVIVQVARNMNFEPFLATLSHD 122

25 Query: 170 APVGMKNPTSGNLKVMFNAVYAQNCQELFYQNKQVKTGDNLLSHVIL--RGYNADYRS 227
 P+ K + A+ Q+ + S++IL RG D
 Sbjct: 123 KPILFKRGLMATIEYLLGALSVLQDTGK-----SNIIICRGVVRGYD--- 164

30 Query: 228 IPNYHYENLLETTTHYESTULGNFFIVVDTHNDSGQK-FLEQIRIVKSVLAD 279
 + + +++ ++TDL I+VD +H + L +I K+V A+
 Sbjct: 165 VETRNMLDINAVPIIQCKTDL---IIVDVSHSTGRDILLPAKIAKAVGAN 214

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1167

35 A DNA sequence (GBSx1243) was identified in *S.galactiae* <SEQ ID 3625> which encodes the amino acid sequence <SEQ ID 3626>. This protein is predicted to be AcpS (acpS). Analysis of this protein sequence reveals the following:

40 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3620 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG22706 GB:AF276617 acyl carrier protein synthase; AcpS
 [Streptococcus pneumoniae]
 Identities = 61/117 (52%), Positives = 90/117 (76%), Gaps = 1/117 (0%)

50 Query: 1 MIVGHGIDLQIEIAITKAYERNRPFARRVLTEQELLFPGKISNPFQMSFLATGRNAAKEA 60
 MIVGHGID+ +E +I A R+ EA+RVLT CE+ F + +RQ+ +L GRW+AKEA
 Sbjct: 1 MIVGHGIDIEELASIESAVTRHEGFAKRVLTAEQMRERTSLK3-BRQIEYLAGRWASAJEA 59

55 Query: 61 YSKALGTGIGKVNFDHDIILSDKGAPLITKEPFGKSPVSIHSGNYAQAQSVILEE 117
 +SKA+GTGI K+ F D+E+L++++GAP ++ PF+GK ++SISH+ + ASVILEE
 Sbjct: 60 PSKAMGTGISKLGPDQLEVINNRGAPYFSQAPFSGKIWLISHTDQVFTASVILEE 116

-1308-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3627> which encodes the amino acid sequence <SEQ ID 3628>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2001 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 76/119 (63%), Positives = 99/119 (82%), Gaps = 1/119 (0%)

Query: 1 MIVHGIDLQIEIAITKAYENRQFAERVLTEQELLFKGISNPKRQMSFLTRWAAKEA 60
MIVHGIDLQIEI AI K Y+RN RFA+++LTECEL +F+ KR+++L GRW+ KEA
Sbjct: 1 MIVHGIDLQIEISAIEKVYQNRFRPAQKILTEQELAIPEFPFY-KRRLNYLAGRWGKEA 59

Query: 61 YSKALGTGIGKVNPHDIELSDKGAFLITKEPFNGKSFVSIHSGNYAQASVILEEEK 119
++K+GIGIG+ F DIELL+D +G P++TK PF G SP+SISHSGNY QASVILE++K
Sbjct: 60 FAGALGTGIGRLTFQDIELINDVKGCPILTSPFKNGSPFISISHSGNYQASVILEDKK 118
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1168

A DNA sequence (GBSx1244) was identified in *S.agalactiae* <SEQ ID 3629> which encodes the amino acid sequence <SEQ ID 3630>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
INTEGRAL Ldkelihood = -3.24 Transmembrane 78 - 94 ( 77 - 97)

----- Final Results -----
bacterial membrane --- Certainty=0.2296 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD51027 GB:AF171873 alanine racemase [Streptococcus pneumoniae]
Identities = 227/366 (62%), Positives = 270/366 (73%)

Query: 1 MISAHRPTRALIDLEAIANNVKSQVSHIPSDDKTPAVVKANAYGHGAVEVSKYIESVD 60
M +S HRPT+ALI L AI N++ + HIP AVVKANAYGHGAV V+K I+ VD
Sbjct: 1 MKASHRPTRALIHGAIQRNQQGAHPQGTLLKLVVKANAYGHGAVAKAIQDQVD 60

Query: 61 GFCVSNLDEAIEIRQAGIVKMLVLGVMPQEVILAKNENITLTVASLEWLRLQISAVD 120
GFCVSN+DEAIEIRQAG+ K IL+GV E V LAK + TLTVL LSW++ VD
Sbjct: 61 GFCVSNIDEAIEIRQGLSKPILILGVSEISAVALAKRYDPLTLTVAGLEWQALLKEVD 120

Query: 121 LSGLEVHIKVDSGMRIGRVQLDGNKILISELGSBGSAVKGIPTHFATADKADNCKFNQ 180
L+GL VH+K+D+SGMRIG R+ E + L + G V+GIPTHFATADE + FN Q
Sbjct: 121 LTGLTVILKIDSGMRIGFRFASVEBQAQDQLQHGVCVBEIPTHFATADESDYFNQA 180

Query: 181 LTFPKDFISGLDNCPLVHASNSATSLWHSSETIFNAVRGVVMYGLNPSGTDLDLPYIN 240
L FK ++ + P+LVHASNSAT+LWH ETIFNAVR+G MYGLNPSG LDLPY +
Sbjct: 181 LERFKTILASMKVPELVHASNSATTLMHVSTIFNAVRKIDAMYGLNPSGAVLDLPYIL 240

Query: 241 PALSLSESLAVHVKQLHDGSGVQVGYATQVIGDEFGVTGVPYIGYADGWRDMQGSFVINGE 300
PAL+LES LVHVK + G +GVGATQ ++ + TVPIGYADGWRDMQ FSV+V+G+
Sbjct: 241 PAITLESALAVHKTVPAGACMGVGYATQADSEQVIATVPYIGYADGWRDMQNSFVLVNGQ 300
```

-1309-

Query: 301 LCEIIGRVSMQDQITIRLPKQYTGKVTILIGQQGSCNITTTDVAQKQRTINYEVCCLSD 360
 C I+GRVSMQD+TIRLP+ Y+GKVTILIG G IT T VA R TINYEV+CLSD
 Sbjct: 301 ACPIVGRVSMQDQITIRLPKLYPLGKVTILIGSNGDKKITATQVATYRVVTINYEVCCLSD 360

5 Query: 361 RIFRY 366
 RIFR Y
 Sbjct: 361 RIFREY 366

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3631> which encodes the amino acid sequence <SEQ ID 3632>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.34 Transmembrane 82 - 98 (82 - 98)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1935 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:AADS1027 GB:AF171873 alanine racemase [Streptococcus pneumoniae]
 Identities = 222/366 (60%), Positives = 273/366 (73%)

25 Query: 1 MISSFHRPTVARVNLQAIKENVASVQKHIPLGKVTYAVVKADAYGHGAVQSKALLPQVD 60
 M+S HRP T A ++L AI++N+ + HIP G AVVKA+AYGHGAV V+KA+ VD
 Sbjct: 1 MKASFHRPTKALHLAGIRQNIQCGAHIPQSTLKLAVVKANAYGHGAVAKAIQDDVD 60

Query: 61 GYCNSNLEALQLRQAGIDKKEILLGLVLLPNELELVANAITVTIASLWIALARLEKKE 120
 G+CVSNLEA++LRQAG+ K ILILGV + LA T+T+A L+WI ++ +
 30 Sbjct: 61 GFCVSNIDEA+RLRQAGLSKPIILIGVSRIRALAKAYDPTLTVAGLEWIALQALLDEVD 120

Query: 121 CQGLKVHVKKVDSGMRIGLRSGKEVNLLIDSLKELGADVGEIPTHFATADEADTKFNQ 180
 GL VH+K+DSGMRIG R + EV D L++ G VEGIFTHFATADE D FN Q
 35 Sbjct: 121 LTGLTVHLKIDSGMRIGFREASEVEQQLLQHQGVCEGIFTHFATADESDDYFNAQ 180

Query: 181 LQFFKLLAGLEDKPLVHASNSATSIWHSOTIPNAVRLGIVSYGLNPSGSDLSLPPFLQ 240
 L+ FK ++A +++ P LVHASNSAT++WH +TIFNAV+G YGLNPSG+ L LP+ L
 Sbjct: 181 LERFKTILASMKVEPELVHASNSATTLWHVETIPNAVNGDAMYGLNPSGAVLDEYDLI 240

40 Query: 241 EALGLESLVHVKMSAGDTVGKATYTTAKKSEYVGTVPVIGYADGWIRNMOGFSVLVDGQ 300
 AL+LES+LVHVK + AG +GYGATY A + + TVPIGYADGWIR+MQ FSVLVDGQ
 Sbjct: 241 PALTLESALVHVKTVPAGACMGYATQADSEQVIATVPVIGYADGWIRNMQFSLVDGQ 300

45 Query: 301 FCEIIGRVSMQDQITIRLPKAYPLGKVTILIGSNQQKQHNISTDIANYRNTINYEVCCLSD 360
 C I+GRVSMQD+TIRLPK YELGKVTILIGS N I+ T +A YR TINYEV+CLSD
 Sbjct: 301 ACPIVGRVSMQDQITIRLPKLYPLGKVTILIGSNGDKKITATQVATYRVVTINYEVCCLSD 360

Query: 361 RIFRY 366
 RIFR Y
 50 Sbjct: 361 RIFREY 366

An alignment of the GAS and GBS proteins is shown below.

Identities = 247/366 (67%), Positives = 295/366 (80%)

55 Query: 1 MISSFHRPTRALIDLEAIANVSVQKHIPSDKKTFAVVKANAYGHGAVEVSKYIESVD 60
 MISS+HRPT A ++L AI NV SVQ+HIP KT+AVVKA+AYGHGAV+VSK + VD
 Sbjct: 1 MISSFHRPTVARVNLQAIKENVASVQKHIPLGKVTYAVVKADAYGHGAVQVSKALLPQVD 60

60 Query: 61 GFCVSNLEAIELRQAGIVKMLLVGVVMPVEQVILAKNENITLTVASLEWIALCQTSYVD 120
 G+CVSNLEA++LRQAGI K IL+GV++P ++ IA IT+T+ASL+H+ L + +
 Sbjct: 61 GYCNSNLEALQLRQAGIDKKEILLGLVLLPNELELVANAITVTIASLWIALARLEKKE 120

Query: 121 LQGLVHVHKVDSGMRIGVRQGLDEGNKLSLSELGSGASGVKGIPTHFATADEADNCKFNQ 180
 GL+VH+KVDGMRIG+R E N LI L E GA V+GIFTHFATADEAD+ KFNQ

-1310-

Sbjct: 121 CQGLKVHVKVDSGMRIGLSSKEVNLIDSLKELGADVGGIPTHFATDEADDTKFNQQ 180
 Query: 181 LTFPKDFISGLDNCPLVHASNGATSLWHSSTIPNAVRLGVVMYGLAPSGTDLDPVPTN 240
 L PFK I+GL+ P LVHASNGATS+WHS+TIPNAVRLG+V YCLNPSG+DL LP+L+
 5 Sbjct: 181 LQFFKKLIAGLEDKPLVHASNGATSINWSDTIPNAVRLGIVSYGLNPSGDLDPFPLQ 240
 Query: 241 PALSLSSSLVHVKQLHDGSGVGVGATYQVVGDEFVGVTVPIGVADGNTRMDQGSFVIVNKE 300
 ALSLES LVHK + G VGVGATY E+VGVTVPIGVADGNTR+MQGSFV+V+G+
 10 Sbjct: 241 KALSLESLSLVHVKMTSAGTVGVGATY+AKKSEYVGVTVPIGVADGNTRNMQGSFVIVNDQ 300
 Query: 301 LCBIIGRVSMDDQHTIRLPQKTYIGTKVTLIGQQGSCHIT+TDVAKRQTINYEVLCLLSD 360
 CBIIGRVSMDDQ+TIRL+ Y +GTYKVLIG NI+TD+A R TINYEVLCLLSD
 Sbjct: 301 FCBIIIGRVSMDDQLTIRLPKAYPLGTKYVLIGSNQKMIISTDIANYRNTINYEVLCLLSD 360
 15 Query: 361 RIRPYR 366
 RIRP Y
 Sbjct: 361 RIRPIY 366

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1169

A DNA sequence (GBSx1245) was identified in *S. agalactiae* <SEQ ID 3633> which encodes the amino acid sequence <SEQ ID 3634>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

25 Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 30 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

There is also homology to SEQ ID 1988.

A related GBS gene <SEQ ID 8745> and protein <SEQ ID 8746> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 4
 McG: Discrim Score: 8.81
 GVH: Signal Score (-7.5): 0.659999
 Possible site: 27
 40 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 1.06 threshold: 0.0
 PERIPHERAL Likelihood = 1.06 247
 modified ALOM score: -0.71
 45 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear)

SEQ ID 8746 (GBS98) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 5; MW 80kDa).

GBS98-His was purified as shown in Figure 192, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1170

A DNA sequence (GBSx1246) was identified in *S. agalactiae* <SEQ ID 3635> which encodes the amino acid sequence <SEQ ID 3636>. This protein is predicted to be junction specific DNA helicase (mmsA) (recG). Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.16 Transmembrane 530 - 546 (530 - 546)

----- Final Results -----
bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA90280 GB:Z49988 MmsA [Streptococcus pneumoniae]
Identities = 483/671 (71%), Positives = 568/671 (83%)

Query: 1 MLLQSPISNLKGGPKSAEKFKLDIYTVEDLLLYPPRYEDFKSKSVFDLVGGEKAVIT 60
N L P+ L G GKPSAEK+ KL I ++DLLLY+PPRYEDFK+K V +L DGEKAV++
Sbjct: 1 MNLHQPLHLVLFVGPKSAEKYAKLGIEINQDOLLNFFPRYEDFKTKQVLEDEGEKAVLS 60

Query: 61 GLVVTPANVCYVGKRNRLSFLRQGEAVLNVSFNPYADKIELQGEVAVFGKWDATK 120
G VVTPA+VQYVGKRNRL F L+QGE V V+FPNQPYLADKIELG +AVFGKWD K
Sbjct: 61 GQVVT PASVCYVGKRNRLSFLRQGEVVFVNFNFCYPLADKIELGATLAVFGKWDRAK 120

Query: 121 SAITGMKVLQAVEDDQPVYRVAQGISQSTLTKAIKSAPEISAHLEKENLPATLLEKYR 180
++TGMKVLQAVEDD+QPVYR+AQGISQ++L+K I K+AF+ L ++ENLP +L+KRY+
Sbjct: 121 ASLTGMKVLQAVEDDQPVYRLAQGISQASLVKVIKTAFQSLDLLEENLPQSLDLKYK 180

Query: 181 LMGRSQACLAMEFFPKDITSEYKQALRRIKFEELFYFQNLQVLKSENKSENGLPILYKSH 240
LM R QA AMHFFPKD+ EYKQALRRIKF EL+YFQM LQ LKSEN+ + +GL + +S+
Sbjct: 181 LMSRQAVRAMHFFPKDLAEYKQALRRIKFAELFYFQMLQLKSENVRQSGSLVNLMSQ 240

Query: 241 AMETKISSLPFILTNAQKRLDEILSDMSSGAHNRLLQGDVGSGKTVIAGLSMYAAYTA 300
+ +SLPF LT AQ++GL EIL+DM S HNRLLQGDVGSGKTV+AGL+M+AA TA
Sbjct: 241 KVTAVKASLPFALTQAQKSLQELITDMKSDHNRLLQGDVGSGKTVVAGLMAFAAYTA 300

Query: 301 GPQSALMVPTTEILAEQHYSIQELFPDLISAIITSGMKAAVKRTVLAAIANGSDVMIVGT 360
G+Q+ALMVPTTEILAEQH+ SLQ LPP+L +A+LT +KAA KR VL I A G D+I+GT
Sbjct: 301 GYQALMVPTTEILAEQHESFELQNLFPNLKALLTSGSKAAEKREVLFTAKGEADLIIGT 360

Query: 361 HALIQDSVCYHKLGLVITDQRRHFGVKQRRIFREKGENEDVLAMTATPIPTLTAITAFGE 420
HALIQD V+Y +LGL+I DEQRHFGV QRRI REKG+NEVDLAMTATPIPTLTAITAFG+
Sbjct: 361 HALIQDGVGYARILGLIIDIQRHFGVQRRIILREKGENEDVLAMTATPIPTLTAITAFG 420

Query: 421 MDVSIIDELPAGRKPIITRWIKHQLQTLVSWKSELQDAQVVISPLIESEALDLGN 480
MDVSIID++PAGRKPI+TRW+KHEQL VL W++GE+QK +Q YVISPLIESEALDLGN
Sbjct: 421 MDVSIIDQMPAGRKPIITRWIKHQLQVTLWKEIQRISQAYVISPLIESEALDLGN 480

Query: 481 AVALHAEISLTYFEGIAKVALVHGRNKNDEKDAIMQDPKDKSHILLVSTTVIEGVNVNVA 540
A+AL EL+T+F G A+VAL+HGRNK+DEK IMQDPK+K+ ILLVSTTVIEGVNVNVA
Sbjct: 481 AIALSEELTIFAGKAEVALVHGRNKNSEKDKIMQDPKERTQILVSTTVIEGVNVNVA 540

Query: 541 TIMIMDADRFGLSQLHQLRGRVGRGKQSYAVLVANPSTDGSKRMTIMETTTGFLA 600
T+MIMDADRFGLSQ+HQLRGRVGRG KQSYAVLVANPSTDGSK RM IMETTT+GFLA
Sbjct: 541 TVMIMDADRFGLSQLHQLRGRVGRGKQSYAVLVANPSTDGSKRMTIMETTTGFLA 600

Query: 601 ESDLAMRGSGETFSTRQSGIPFQVADIEDYPILEBARVASDIVKDNKNKENTWALI 660
E DLAMRGSGETFSTRQSG+PEFQVADI+ED+PILEBAR+VAS I W+E EW +I
Sbjct: 601 EEDLAMRGSGETFSTRQSGLPFQVADIIEFPILEBARVASYISSLEAWQDEPWMMI 660

Query: 661 LBNLRQHSDFD 671
+L + D
Sbjct: 661 ALHLEKKEHLD 671

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3637> which encodes the amino acid sequence <SEQ ID 3638>. Analysis of this protein sequence reveals the following:

Possible site: 17
>> Seems to have no N-terminal signal sequence
10 INTEGRAL Likelihood = -0.16 Transmembrane 530 - 546 (530 - 546)
----- Final Results -----
bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 641/671 (95%), Positives = 655/671 (97%)

20 Query: 1 NLLQSPISNLKGFPGKSAEKFKQLDIYTVEDLLLYFFRYEDPKSKSVFLVDGKXAVIT 60
N+L +P+SNLKGFGPKSAEKFKQLDIYTVEDLLLYFFRYEDPKSKSVFLVDGKXAVIT
Sbjct: 1 MILTAPMSNLKGFPGKSAEKFKQLDIYTVEDLLLYFFRYEDPKSKSVFLVDGKXAVIT 60

25 Query: 61 GLVVTTPANVQYTGKKNRLSPKLRQGEAVLNVSPFNQPYLADKIELQGEVAVFGKIDATK 120
GLVVTTPANVQYTGKKNRLSPKLRQGEAVLNVSPFNQPYLADKIELQGEVAVFGKIDATK
Sbjct: 61 GLVVTTPANVQYTGKKNRLSPKLRQGEAVLNVSPFNQPYLADKIELQGEVAVFGKIDATK 120

30 Query: 121 SAITGMKVLQVDEDDNQPYVRVAGGISQSTLIKAISAFETSAHLEKLENLPAITLLEKYR 180
SAITGMKVLQVDEDDNQPYVRVAGGISQSTLIKAISAFETSAHLEKLENLPAITLLEKYR
Sbjct: 121 SAITGMKVLQVDEDDNQPYVRVAGGISQSTLIKAISAFETDAHLEKLENLPAITLLEKYR 180

35 Query: 181 LMGSRQACLAMHFPKDIITYKQALRRIKFELFYQMNQLVLAENKSETNGLPILYSKH 240
LMGSRQACLAMHFPKDIITYKQALRRIKFELFYQMNQLVLAENKSETNGLPILYSKH
Sbjct: 181 LMGSRQACLAMHFPKDIITYKQALRRIKFELFYQMNQLVLAENKSETNGLPILYSKH 240

40 Query: 241 AMETKISSLPFILINAQKRSLLDILSDMSSGAHMRLLQGDVSGSKTVIAGLSMYAYTA 300
AMETKISSLPFILINAQKRSLLDILSDMSSGAHMRLLQGDVSGSKTVIAGLSMYAYTA
Sbjct: 241 AMETKISSLPFILINAQKRSLLDILSDMSSGAHMRLLQGDVSGSKTVIAGLSMYAYTA 300

45 Query: 301 GPQSALMVPTTEILAEQHYISLQELFPDLSIAITSGMKAQVKTIVLAIAANGSVDMIVGT 360
GPQSALMVPTTEILAEQHYISLQELFPDLSIAITSGMKAQVKTIVLAIAANGSVDMIVGT
Sbjct: 301 GPQSALMVPTTEILAEQHYISLQELFPDLSIAITSGMKAQVKTIVLAIAANGSVDMIVGT 360

50 Query: 361 HALIQDSVQYHKLGLVITDQHRFGVQKRRIPREKNGFDVLMATATPIPTLTAITAPGE 420
HALIQDSVQYHKLGLVITDQHRFGVQKRRIPREKNGFDVLMATATPIPTLTAITAPGE
Sbjct: 361 HALIQDSVQYHKLGLVITDQHRFGVQKRRIPREKNGFDVLMATATPIPTLTAITAPGE 420

55 Query: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGLQKDAQVVISPLIESSEALDLGN 480
MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGLQKDAQVVISPLIESSEALDLGN
Sbjct: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGLQKDAQVVISPLIESSEALDLGN 480

60 Query: 481 AVALHAEISTYPEGIAKVALVHGRMKNDEKDAIMQDFDKKSHILVSTTVIEGVNVFNA 540
AVALHAEISTYPEGIAKVALVHGRMKNDEKDAIMQDFDKKSHILVSTTVIEGVNVFNA
Sbjct: 481 AVALHAEISTYPEGIAKVALVHGRMKNDEKDAIMQDFDKKSHILVSTTVIEGVNVFNA 540

65 Query: 541 TIMIMDADRFLGSLQLHLGRGVRGKQSVAVLVANPKTDSGKRMITMETTIDGFVLA 600
TIMIMDADRFLGSLQLHLGRGVRGKQSVAVLVANPKTDSGKRMITMETTIDGFVLA
Sbjct: 541 TIMIMDADRFLGSLQLHLGRGVRGKQSVAVLVANPKTDSGKRMITMETTIDGFVLA 600

60 Query: 601 RSDLRGSGSGEIGFTRQSGIPEPCVADIVEDYPILEBARVSDIVKNNKENTENAL 660
RSDLRGSGSGEIGFTRQSGIPEPCVADIVEDYPILEBARVSDIVKNNKENTENAL +V+ IV D NM
Sbjct: 601 RSDLRGSGSGEIGFTRQSGIPEPCVADIVEDYPILEBARVSAIVSDNNKENTENAL 660

Query: 661 LBNLRQHSDFD 671
N+R+ +D

-1313-

Sbjct: 661 AQNIRKKKEVVD 671

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1171

A DNA sequence (GBSx1247) was identified in *S.agalactiae* <SEQ ID 3639> which encodes the amino acid sequence <SEQ ID 3640>. This protein is predicted to be aryl-alcohol dehydrogenase (b1647). Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1562 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10105> which encodes amino acid sequence <SEQ ID 10106> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:BA07646 GB:AP001520 aryl-alcohol dehydrogenase [Bacillus halodurans]
   Identities = 173/300 (57%), Positives = 224/300 (74%)

Query: 7  IQQTGIQATRIALGCMRMSDLKGKQAEVVGTAIDLDLGINFFDHADIYGGGLSELRFDAI 66
      +G + ++ +A+GCMR++ + K+AE V TAL+ G NFFDHADIYGGG E F DAI
25 Sbjct: 6  LGSSSLSEFVVAVGCMRINATSKKEAERFVQTALDQGANIFFDHADIYGGGCEEIFDAI 65

Query: 67  KHLNVNRDKMIQSKOGIRGYPDFPSKEYILSSVDGILRLTGYLDPLILHRPDVLEVP 126
      + R+K+I+QSKOGIRG PDPSEYIL SVDGIL+RL T+YLD L+LHRPD LVLEVP
30 Sbjct: 66  QMTEAVREKIIQSKOGIRGYPDFPSKEYILQSVGDIQLRLKTDYLDVLLHRPDALVEP 125

Query: 127  EEVAZAFTKLRABGKVKHFGVSNQNRPFQMEILLQSYLDEPLAVNQQLSPARTMFDAGLN 186
      EEVAZAF L + GKV+HFGVSNQN Q+ELL+ ++ +P+ NQLQLS + M +G+N
35 Sbjct: 126  EEVAZAFDILLESSGKVRHFGVSNQNPQIELLKKFVRQP+VANQLQLSITNATMISSGN 185

Query: 187  VNMENKASIEHDDGIVDYCRLEKVTIQAWSPFQIDLSRGLFVNHPDYKELNETIAKLAJN 246
      VNM N+++I D ++DYCRLE VTIQ WSPFQ G+P+ + +ELN+ I +LA+
40 Sbjct: 186  VNMENESAINRDSGLDYCRLEHVTIQWSPFQYGFQVFLGNDLFPELNKKIDELAEK 245

Query: 247  YNVSEAIIVIAWILRHPAKQVAIVGSMNPRLKAIKDANDIALTRKEWYDIYRSAGNILE 306
      Y VS+ I IAW+LRHPA MQ ++G+MN RLK KA++I LTR+EWY+IYR+AGNILEP
45 Sbjct: 246  YEVNSTTIAIAWILRHPAKQVIGTMMNLKRLKDCCKASEIRLTREBWEYIYRAAGNILEP 305

```

There is also homology to SEQ ID 780.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1172

A DNA sequence (GBSx1248) was identified in *S.agalactiae* <SEQ ID 3641> which encodes the amino acid sequence <SEQ ID 3642>. This protein is predicted to be shikimate 5-dehydrogenase (aroE) (aroE). Analysis of this protein sequence reveals the following:

```

50 Possible site: 21
   >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-1314-

bacterial cytoplasm --- Certainty=0.0988 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74762 GB:AE000264 putative oxidoreductase [Escherichia coli K12]
 Identities = 114/279 (40%), Positives = 171/279 (60%), Gaps = 3/279 (1%)

Query: 10 LTGLIANPARHSLSPFMAWNTSPQEKDQFAYLTFEVEEGKLTAEVGVVRLGIRGVNVSM 69
 L GL+A P RHSLSP M N + + + Y+ FEV+ A+ G++AL +RG VSM
 10 Sbjct: 9 LIGLWAYPIRHSLSPFMQNKALEKAGLFPTTMAFEVINDSFGAIEGLKALKMRGTGVSM 68

Query: 70 PFKQSVIPLLDLSPQAKLVGAVITVNMQQTGRLVGHMTDGIQCPKALAAQGFSAIKKI 129
 P KQ +D+L+P AKLVGA+NTIVN G R G+ TDG G +A+ GF K K
 15 Sbjct: 69 PNKQLACEYVDELTPAAKLVGAINITVNDGSLR--GYTWDGTGHIRAKIKSGPDFIKGKT 126

Query: 130 ITIAGIGGSGKAVAVQAMGVAEIRLPNRNNSNYDKVIDLSDKIKKQFQIKVVVDYLEN 189
 + + G G+ A+ Q A+EG+ EI+LNR +DK + + + + V V L +
 20 Sbjct: 127 MVLLGAGGASTAIGACQAIKSLKIKLPNRDEFPDKALAFACRVNENTDCVVTIDLAD 186

Query: 190 KTAFKDAIRTSHPYIDATSLGMKPLNYSLNDPFIILTPMLAVVDLVYKPKETALRFV 249
 + AF +A+ + + + T +GM+PLAN SLND +L P L+V + VY P T IL+ +
 25 Sbjct: 187 QQAFALASADILNTGTVGMKPLENLSLNDISLHPGLVTECVYNPHMTKLQQAQ 246

Query: 250 QNGVGHAYNGLMLTYCGAKAPOLITNQMPISSVERVL 288
 Q G K +G GML+QGA E F L T + + P+ V+V+
 25 Sbjct: 247 QAGCK-TIDYGMMLMQGAEQFTLWTGRDFFLEYVKQVM 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3643> which encodes the amino acid
 30 sequence <SEQ ID 3644>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40 >GP:AAC74762 GB:AE000264 putative oxidoreductase [Escherichia coli]
 Identities = 132/280 (47%), Positives = 186/280 (66%), Gaps = 3/280 (1%)

Query: 11 LVSLATPIRHSLSPFMNINEAYALGLDYLAPEVGTQLADAVQIRALGIRGVNVSM 70
 L+ L+A PIHSLSP+M N+A K GL + Y+FEV + A++G++AL +RG+ VSM
 45 Sbjct: 9 LIGLWAYPIRHSLSPFMQNKALEKAGLFPTTMAFEVINDSFGAIEGLKALKMRGTGVSM 68

Query: 71 PNKEAILPLLDLSPAAKLVGAVITVNMKGKHLVGHITDGTGALRALADEGVSVNKKI 130
 PNK+ +D+L+PAA+LVGA+NT+VN DG +L G+ TDG G +R+A+ + G +K K
 50 Sbjct: 69 PNKQLACEYVDELTPAAKLVGAINITVNDG--YLRGYNTDGTGHIRAKIKSGPDFIKGKT 126

Query: 131 ITIAGVGGGKALAVOLAFDGAKSVLENNRQTRLSSVQKLVTKINQLTRTKVTLDQLED 190
 + L G GGA AI Q A + G KE++LNR+ ++N+ T V+ DL D
 55 Sbjct: 127 MVLLGAGGASTAIGACQAIKSLKIKLPNRDEFPDKALAFACRVNENTDCVVTIDLAD 186

Query: 191 QTAFKAIRESHLPIDATSVGMKPLENLSLITDPRIELPAVVDIVYSPAEITKLAFAR 250
 Q AF EA+ + + + T VGMKPLEN SL+ D L+ P L+V + VY+P TKIL A+
 60 Sbjct: 187 QQAFALASADILNTGTVGMKPLENLSLNDISLHPGLVTECVYNPHMTKLQQAQ 246

Query: 251 QHGAKGVINGLGMVLTYCGAEAFKLTQGMVDAIKPLG 290
 Q G K I+G GN+L+QGA E F L TG+D P+ +K +G
 Sbjct: 247 QAGC-KTIDYGMMLMQGAEQFTLWTGRDFFLEYVKQVM 285

An alignment of the GAS and GBS proteins is shown below.

-1315-

Identities = 166/288 (57%), Positives = 221/288 (76%)

Query: 4 LNGETLLTGLLANPARHSLSPIMWNTSFOEKNMMYAYLTFEVEBQKLTREAVRGVRLGIR 63
 I+G TLL L+A P RHSLSP M N ++ + YAYL FEV +L +AV+G+RALGIR
 5 Sbjct: 5 LSGHTLLVSLIATPIRHSLSPKMHNEAYAKGLDYAYLAFVGTREQLADAVQSGIRALGIR 64

Query: 64 GNVNVSMPFQSVIPLLDLSPQAKLVGAVNTVNGGTGRIAGHMTDGGICPKALAAQGF 123
 G NVSMP K+++PLLDLSP A+LWGAVNT+VN+ G G LWGH+TDGIG +ALA +G
 10 Sbjct: 65 GSNVSMFNKAILPLLDLSPAAELVGAVENTVNNKDGKHLVGHITDGGIALRALADEGV 124

Query: 124 SAKNKIITTAGIGGSGKAVAVQAAMBGVAIRLFNMENSGNYDKVLDSDKIKKQFQIKVV 183
 E RKNKIIT+AG+GG+GKA+AVQ A +G B+RLFNR ++ V L K+ + +KV
 Sbjct: 125 SVNKKIITLAVGGAGKAIYVQLAFDGAKEVRLFNQRNTRSSVQKLVTKLQLITRTKVT 184

Query: 184 VDVLNKTAFKDAIRTSHFVIDATSLGMRPLDNYSLINDPRILTFNLVVVDLVKPKETA 243
 + LE++TAFK+AIR SH +IDATS+GM+PL+N SLI DPE++ P+LVV D+VY P ET
 15 Sbjct: 185 LQQLSDQYAFKBAIRESHLFIDATSVGMKPLNLSLITDPELIRPDLVVFDIVYSPTK 244

Query: 244 LIRFVRGNVVKHAYNGLMLIYQGAFAQLITNQEMPISSVERVLQTE 291
 LL F RQ+G + NGLGM+YQGAFAF+LIT Q+MP+ +++ +L E
 20 Sbjct: 245 LLAFARQHGSAQKVLNGLMVLTYQGAFAFLITQDMPVDAIKELLGDE 292

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1173

A DNA sequence (GBSx1249) was identified in *S.agalactiae* <SEQ ID 3645> which encodes the amino acid sequence <SEQ ID 3646>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 30 INTEGRAL Likelihood = -6.16 Transmembrane 57 - 73 (53 - 76)

----- Final Results -----
 bacterial membrane --- Certainty=0.3463 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1174

A DNA sequence (GBSx1250) was identified in *S.agalactiae* <SEQ ID 3647> which encodes the amino acid sequence <SEQ ID 3648>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2333 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10103> which encodes amino acid sequence <SEQ ID 10104> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05343 GB:AP001512 L-asparaginase [Bacillus halodurans]
Identities = 158/319 (49%), Positives = 214/319 (66%), Gaps = 4/319 (1%)

5 Query: 1 MKKILVLTGGTISMRANEKQVMSADNPMKVDLSLDDL-DLAVVDVLNLPSPQITPH 59
MKK+LV+TGGT+M+ +EKG V NP+ SL + + V DFLN+DSP +TP
Sbjct: 1 MKKVLVHTGGTIAHMEDEKGGVCKPKNLPFATVBSLTSASIEVDVDFINLPSPHPTP 60

10 Query: 60 HMLDIHYHLKQHASN--FDGVVTHGTDTLEETAYFLDTMLFKPIITIGAMRSNELG 117
M + LK N FDGVVTHGTDTLEETAY LD ++ ++P++TGAMRS+NELG
Sbjct: 61 LMFQLARLKSRSVGNBSFDGVVTHGTDTLEETAYLLDILLWVEPVVVTGAMRSNELG 120

15 Query: 118 SDGIVNYLTALRVASSDKAKEKGVLVVNMDEIHAQYVTKHTNTLSTFQTPHGLIGI 177
+DG +N++A++ A + +A KGVLVV NDEIH AK VTKHT+NV+TFQ+P +GP+GI+
Sbjct: 121 ADGPHNFIASVKTATDEAKGKGVLVVNMDEIHTAKNVTKHTNSVATFQSPQGPVIGI 180

20 Query: 178 MKQDLLFFKATEERVRFPDLKITGTVPVIVKAYAGKMGDSGIISPLNSQISGLVIEALG 237
K + F A + + + I V ++KAYAGM D ++ + I GLVIEA G G
Sbjct: 161 TKRGVTHPHAPSYKESYTVSSIDHRVVLKAYAGM-DGSVVDALADTGDGLVIEAFQG 239

25 Query: 238 NMPPKTAQETIEELIEQGVFVVLVSRFCNGIABFVYGYEGGAKLQBSGVFMFKELNAPKA 297
N+PP I+L + +PVVLVSR +GI + Y YEGGG L++GV+F LN KA
Sbjct: 240 NLPFAVVPVSIKRLHQAQNFVVLVSRVSGIVQETVAYEGGGRIKLGLGVIFTNGLNGQA 299

25 Query: 298 RLKLLIALNAGLTGQNLKD 316
RLKLL+AL +L++
Sbjct: 300 RLKLLVALELTDRKKLQE 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3649> which encodes the amino acid sequence <SEQ ID 3650>. Analysis of this protein sequence reveals the following:

Possible site: 16
>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.28 Transmembrane 245 - 261 (243 - 261)

35 ----- Final Results -----
bacterial membrane --- Certainty=0.1914 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA05343 GB:AP001512 L-asparaginase [Bacillus halodurans]
Identities = 158/320 (49%), Positives = 218/320 (67%), Gaps = 5/320 (1%)

40 Query: 1 MKKILVLTGGTISMQADNSGRVVPQNDNM-TKIHAAQDQICLVSDFLNLPSPHITPH 59
MKK+LV+TGGT+M D G V P + NP+ + + + V DFLN+PSPH+TP
Sbjct: 1 MKKVLVHTGGTIAHMEDEKGGVCKPKNLPFATVBSLTSASIEVDVDFINLPSPHPTP 60

50 Query: 60 HMLSIYHIQERT--DVFDGIVTHGTDTLEETAYFLDTMALFNIPVVLVIGAMRSSNEV 117
M + ++ R + FDG+VTHGTDTLEETAY LD + L +PVV+TGAMRSSNE+
Sbjct: 61 LMFQLARLKSRSVGNBSFDGVVTHGTDTLEETAYLLDILLWVEPVVVTGAMRSSNEL 119

55 Query: 118 GSDGIYNYLTALRVASSDKAKEKGVLVVNMDEIHAQYVTKHTNTLSTFQTPHGLIGI 177
G+DG +N+++A++ A++D+AK KGVLVV NDEIH AK VTKHT+NV+TFQ+P +GP+GI
Sbjct: 120 ADGPHNFIASVKTATDEAKGKGVLVVNMDEIHTAKNVTKHTNSVATFQSPQGPVIGI 179

60 Query: 178 IMKNDLLFFKTAERPRIFDLRCISGTIPIIKAYAGKMGDSGISLILTPGSIQGLVIEALGA 237
+ K + F + + + I + ++KAYAGM DGS++ + I GLVIEA G
Sbjct: 160 TKRGVTHPHAPSYKESYTVSSIDHRVVLKAYAGM-DGSVVDALADTGDGLVIEAFQG 238

60 Query: 238 GNVPLAVGRIEHLIALGIFVLVSRFCNGIABFVYAYEGGAKLQBSGVFMFKELNAPKA 297
GN+PP V I+L IFV+LVSR +G+ + YAYEGGG L++GV+F LN K
Sbjct: 239 GNLPAVVPVSIKRLHQAQNFVVLVSRVSGIVQETVAYEGGGRIKLGLGVIFTNGLNGQA 298

Query: 298 ARKLLIALNAGLTGQNLKD 317

-1317-

ARLKL+AL ++L++
 Sbjct: 299 ARLKLVALELTTDRKKLQE 318

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 242/321 (75%), Positives = 275/321 (85%), Gaps = 1/321 (0%)

Query: 1 MKKILVLTGTTISMNANEKQVMSSADNPMKYVDLSLDDDLTVVDPLNLSPQITPHH 60
 MKKILVLTGTTISM A+ G+V+ + DNP + + D+ LTV DPLNLSP ITPH
 Sbjct: 1 MKKILVLTGTTISMQADNSGRVVPNDNPMKIHAAQDIQLTVSDPLNLSPHITPHH 60

10 Query: 61 MLDIYHYLKQHASNFDGVTHTGTDITLESTAYFLDTMILP-KIPIITGAMRSNTELGS 119
 ML IYH+++ FDG+VITHGTDITLESTAYFLDTM LP IP+++TGAMRS+NE+GSD
 Sbjct: 61 MLIYTHIEIKERTDVFQGIIVITHGTDITLESTAYFLDTMALPTNI PVVLTGAMRSSNEVGSD 120

15 Query: 120 GVNLYLSALRVANSTOAKDGLVVMNDEIHAAKYVTKIHTNVSTPQTPTGFLGIIMK 179
 G+YNYL+ALRVA+S KA +KGVLVMNDEIHAAKYVTKIHTN+STPQTPTGFLGIIMK
 Sbjct: 121 GIYNYLTALKVASSDKAKKGLVVMNDEIHAAKYVTKIHTNISTPQTPTGFLGIIMK 180

20 Query: 180 QDLFFKATSEVRVFLDLKITGTVPVIVKAYAGMGDSGIISFLNSQMSGLVIEALGAGNM 239
 DIAFFK E R+RFDL I+GT+PI+KAYAGMD I+S L +I GLVIRAIAGAGN+
 Sbjct: 181 NDLLFFKTABPIRFDLRCISGTIPIIKAYAGMGDSGISLLTTPGSIQGLVIEALGAGNV 240

Query: 240 PPKAAQSEIEELKCGVFFVLVSRCFNGIAEPVYTGEGGAKLQSSGVMPVKELNAPKARL 299
 PP A EIE LI G+PV+LVSRCFNG+AEFVY YEGGGA LQE+GVMPVKELNAPKARL
 Sbjct: 241 PPLAVGEIEHLIALGIPVLVSRCFNGMAEPVYATYEGGAMQLQAGVMPVKELNAPKARL 300

25 Query: 300 KLLIALNAGLTQQLADYIEG 320
 KLLIALNAGLTQ LADYIEG
 Sbjct: 301 KLLIALNAGLTQQLADYIEG 321

30

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1175

A DNA sequence (GBSx1251) was identified in *S. agalactiae* <SEQ ID 3651> which encodes the amino acid sequence <SEQ ID 3652>. Analysis of this protein sequence reveals the following:

35 Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4427 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB85142 GB:AL162757 conserved hypothetical protein [Neisseria meningitidis Z2491]
 Identities = 87/285 (30%), Positives = 138/285 (47%), Gaps = 35/285 (12%)

50 Query: 4 KAVFFDIDGTLINDRKNVQKSTIK-AIRNLKDQGIIVGLATGRG---PSFVQFFLENLG 58
 K VFFDID TL + + ++K A+ L+ +GIL LATGR P V+ + G
 Sbjct: 11 KIVFFDIDDTLYRKYTDTLRSVKTAAALRGKILTLATATGRSLATPEKVRDMMASTG 70

Query: 59 LDFAVTYNGQYIYSRSEIYVITQLSKTIVYRLIRYAGARRRKEISLTASGLLGGSGIIGLG 118
 +D VT NQG+ + + + R+ + SLG +G G+
 Sbjct: 71 MDAVVTINGQFALLHGKTVCVEPMDAGIMGRVCAHLD-----SLQMDYAFVGGG--GIA 122

55 Query: 119 TSLRGQIVSSLVPRKAKAIERSPKHFIRRIKPNIDSLMVLRRPIYQVVLVATEGE-- 176
 S L + V R+ KH I ++YQ+++A E E
 Sbjct: 123 VSALSBECVC-----RALKH-----IASDFFADKDYFSKFPVQMLVFAENEMP 166

60 Query: 177 --SERIQKQPPRVKLTSSPSYMDVISEBQSKVKGIERVGQRYGDFSEVIAFGDSNDI 234

-1318-

S+ +++++ +K R ++D++ G SK GI V + G +++V+AFGD ND+
 Sbjct: 167 LWSDIVRE--GLTIVRHESAVDLLPAGASKTDGIRSVVHALGLEMDVWAPGDGLNDV 224
 Query: 235 EMLSGVIGVAMGNASQOVREARYTTADNNDGISKALAHYGLI 279
 EMLSAVG GVAMGN Q +E A+Y ++DG+ L G+I
 Sbjct: 225 EMLSEVGVGVAMGNGBQAKAEAKYVCPGVDEDSVLRGLQDLGV 269

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3653> which encodes the amino acid sequence <SEQ ID 3654>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6014(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 320/459 (69%), Positives = 391/459 (84%)
 Query: 1 NAIKAVFPDIDGTLINDRKNVQKSTIKAIRNLADQGLWGLATGRGSPVQPFLENLGLD 60
 + +NAVFPDIDGTLINDRKN+QK+T KAI+ LK QGI+VGLATGRGP FVQPFLEN GLD
 Sbjct: 1 LTVKAVFPDIDGTLINDRKNVQKSTIKAIQQLAKQGINVGLATGRGSPVQPFLENLGLD 60
 Query: 61 FAVTTNGQVIYSESEIIYTNQLSKITTYYRLIRYAGARRREISLGTASGLLGSGIIGLGT 120
 FAVTTNGQVI +R +++Y NQL K+ +Y++IRYA ++REISLGTASGL GS II +GTS
 Sbjct: 61 FAVTTNGQVILTRDKVLYQNQLPKSMYKVRIRYANEKRYREISLGTASGLASGRIDMGTS 120
 Query: 121 RLQGVSSLVPRKWKAIERSFKHPIRRIRKQPNISLAVILREPIYQVVLVATGESSRI 180
 GQ+SS VP+ W+ +E SFKH IRRIRKQ+ +L+ I+REPIYQVVLW+ +E++I
 Sbjct: 121 PFGQVISSPVFKSWARTVBGSPKHLIRRIKQSPFNLTVMREPIYQVVLVNASQASTYKI 180
 Query: 181 QKQFPRVKLTRSSPYSDMVISEQSGVKVGIERRVGGVFDLSEVIARGDSNDIMELQGV 240
 Q++P +K+TRSSPY+D+IS QSK+RGIER+G+ +GFLSEV+ARGDSND+EMLS V
 Sbjct: 181 QKQFPHIKITRSSPYSLDLISVDQSKIKGIERLGEMPGFDLSEVMAFGDSNDIMELSGV 240
 Query: 241 GIGVAMGNASQOVREARYTTADNNDGISKALAHYGLIQFEIKTFSSRDENFNKVKSF 300
 GIG+AMGNA V++ A +TT NN+DGISKALAHYGLI P+IEK+P SRDENFNKVKF P
 Sbjct: 241 GIGIAGNAETVVDGAHFTDNNNDGISKALAHYGLIHFDIEKSPKSRDENFNKVKDF 300
 Query: 301 HLMADGETIETPRLYDSKAGRSPDFKVEIEVEFLYAASQGNQKVPDQSRHMLHLDIA 360
 H IMD +TIETPR Y EAG+RS FKVEIEVEFLYAAS+G+Q+ P Q+I +IH A+D+A
 Sbjct: 301 HRLMDSUTIETPRSYTISAGYRSRPFKVEIEVEFLYAASGQDQQTQAIPLHLAGVDQA 360
 Query: 361 RDKVISKDHPEPIAFGEVDALDILLYLTYSFVLGVDPKPLDTVHEANMKIFPDGKA 420
 +KV +K H ETL+G+VDAL DILY TYGSFVLGVDP+P+P+ VHEANM KIFPDGKA
 Sbjct: 361 ANKVQAKGHVETPLIQVDALADLLYPTYSFVLGVDPQPIPEVHEANMAKIFPDGKA 420
 Query: 421 HFDPVTHKILKPDOWREHFAPEPSIRRELDQIQKSLNR 459
 HFDPVTHKI KPD W+E APE +I+RLD Q+QKSL R
 Sbjct: 421 HFDPVTHKIQKPDWQREHFAPEVAIKKELQKQKSLQR 459

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1176

A DNA sequence (GBSx1252) was identified in *S.agalactiae* <SEQ ID 3655> which encodes the amino acid sequence <SEQ ID 3656>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

-1319-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1671 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10101> which encodes amino acid sequence <SEQ ID 10102> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06903 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 61/141 (43%), Positives = 92/141 (64%)

Query: 22 YERILVAIDGSTESSELAPEKAVNVALRNDSEILITHVIDTRALQSPATFDITYIEKLEKE 81
 Y ILVA+DGST+++ A KA N A ++L + HVID+R+ + +D + E +
 Sbjct: 2 YNHILVAVDGSTQAKRALYKAFNYAKEFKADLFICHVIDSRSPATVEQVDRTVGGAELD 61

Query: 82 AKDVLREYEQAREKGADKVRQVIEFGNPKTLAHDIPPEKVDLIMVGATGLNTPERFK 141
 K +L+ Y ++A + G DKV ++FG+PK ++ I +K +DLI+ GATGLN RRF
 Sbjct: 62 GKLLQRYSEBAEKAGVDKVTHTLDPGSPKANISKTIAQYIDILITGATGLNAVERFL 121

Query: 142 IGSSSEYILRHAKVDLLIVRD 162
 +GS SE + RHAK D+LIVR+
 Sbjct: 122 MGSVESVARHAKCDVLIVRN 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3657> which encodes the amino acid sequence <SEQ ID 3658>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1296 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/156 (75%), Positives = 135/156 (86%)

Query: 12 LEEDRLMSQYERILVAIDGSTESSELAPEKAVNVALRNDSEILITHVIDTRALQSPATFD 71
 L+ED MS KY+RILVAIDGS SEELAF K VNVALEND+ L+L HVIDTRALQS ATFD
 Sbjct: 25 LKEDSSMSLYKRILVAIDGSYSELAPEKAVNVALENDATLLVHVIDTRALQSPATFD 84

Query: 72 TYIEKLEKEAKDVLREYEQAREKGADKVRQVIEFGNPKTLAHDIPPEKVDLIMVGA 131
 TYIEKLE+EKADVL+++EKQA+ G ++Q+IEFGNPK LLAHDIP++E DLIMVGA
 Sbjct: 85 TYIEKLEKEAKDVLDDPEKQKQIAGITNIKIQIEFGNPKNLAAHDIPDREANLIMVGA 144

Query: 132 TGLNTPERFKIGSSSEYILRHAKVDLLIVRDPNKM 167
 TGLNTPER IGSSSEYI+RHAK+DL+VRD KT+
 Sbjct: 145 TGLNTPERLLIGSSSEYIMRHAKIDLLVVRDSTKTL 180

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1177

A DNA sequence (GBSx1253) was identified in *S.agalactiae* <SEQ ID 3659> which encodes the amino acid sequence <SEQ ID 3660>. This protein is predicted to be aspartate aminotransferase (aspC). Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

-1320-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2803 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21948 GB:U32714 aminotransferase [Haemophilus influenzae Rd]
 Identities = 142/212 (66%), Positives = 181/212 (84%), Gaps = 1/212 (0%)

Query: 1 MKIFDKSMKLEHVAYDIRGVPVLEADRM RANGEKILRLNTGNPAAGFEAPDEVIRDLIT 60
 M+ F K S KLEHV YDIRGPV +EA R+ G KIL+LN GNPA PGFEAPDE++ D++
 Sbjct: 1 MRLFPKSKLEHVCYDIRGVPVHKEALRLSEGNKILRLNTGNPAAGFEAPDEILVDLRL 60

Query: 61 NARESGYSDSGKIFSAKAVMYYQLQNI-HVMDIYIVNGVSEGISMSMQALLNDD 119
 N ++GY DSKG+SARKA++QTYQ + I ++D+YI NGVSE IS+M+MQALL++ D
 Sbjct: 61 NLPASQGYCDSKGLYSARKAIVQYQTSQKGLGATVNDVYI NGVSELITMAMQALLNDD 120

Query: 120 EVLVPMPOYPLMTACVSLAGNAVHYICDEANWYFDIDDIKSKITKTAIVLINPNP 179
 EVLVPMPOYPLMTA V+L+GG AVHY+CDG+ANW+P IDDIK+K+ +KTKAIV+INPNP
 Sbjct: 121 EVLVPMPOYPLMTAAVTLSSGKAVHYLCDEANWFPTIDDIKAKVNARTKAIVINPNP 180

Query: 180 TGAVYPRSLQEIVDIARQNLLIIPSDVEYDR 211
 TGAVY +E+LQEIV+IARQN+LIIF+DE+YD+
 Sbjct: 181 TGAVYSKELQEIVETARQNLLIIFADEITYDK 212

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3661> which encodes the amino acid sequence <SEQ ID 3662>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2936 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/212 (80%), Positives = 193/212 (90%), Gaps = 1/212 (0%)

Query: 1 MKIFDKSMKLEHVAYDIRGVPVLEADRM RANGEKILRLNTGNPAAGFEAPDEVIRDLIT 60
 MKI +K S KLEHVAYDIRGVPV+EA+RM A+GEKILRLNTGNPAAGFEAPDEVIRDLI
 Sbjct: 13 MKLIEKSKLEHVAYDIRGVPVLEADRMRTASGEKILRLNTGNPAAGFEAPDEVIRDLIV 72

Query: 61 NARESGYSDSGKIFSAKAVMYYQLQNI-HVMDIYIVNGVSEGISMSMQALLNDD 119
 NAR SSGYSDSGKIFSAKAVMYY QL+ VD++DIY+ NGVSE IS+S+QALLND
 Sbjct: 73 NARLSGYSYSDSGKIFSAKAVMYYQLNGPVPVDIEDIYLGNGVSELISISIQALLNDD 132

Query: 120 EVLVPMPOYPLMTACVSLAGNAVHYICDEANWYFDIDDIKSKITKTAIVLINPNP 179
 EVLVPMPOYPLMTACVSL G AVHY+CDG+ANWYFDI DIKSKITS+TKAIV+INPNP
 Sbjct: 133 EVLVPMPOYPLMTACVSLGGKAVHYLCDEANWFYFDIADIKSKITKTAIVLINPNP 192

Query: 180 TGAVYPRSLQEIVDIARQNLLIIPSDVEYDR 211
 TGA+YR+EIL++IV +AR++ LIIF+DE+YDR
 Sbjct: 193 TGAYPEILIEDIVARSHQLIIFADEITYDR 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1178

A DNA sequence (GBSx1254) was identified in *S.agalactiae* <SEQ ID 3663> which encodes the amino acid sequence <SEQ ID 3664>. Analysis of this protein sequence reveals the following:

-1321-

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-14.75 Transmembrane 38 - 54 (29 - 60)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.6901(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 10 A related GBS nucleic acid sequence <SEQ ID 9389> which encodes amino acid sequence <SEQ ID 9390> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3665> which encodes the amino acid sequence <SEQ ID 3666>. Analysis of this protein sequence reveals the following:

15 Possible site: 43

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-15.97 Transmembrane 35 - 51 (25 - 58)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.7389(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 51/87 (58%), Positives = 63/87 (71%), Gaps = 7/87 (8%)

Query: 1 MAKKEWEEKVVENNSHERKDKITRTSRGVVSSSTPWITAFLSGAFFVIVVAILFIVFYTSNKG 60
 MAK+PWE+K+V++ + TR SR STPW+TA LS FFVI+VAILFI FYTSN G
 30 Sbjct: 1 MAKEWEEKIKVDOTIGTR---TRKSRNAFISTPMLTALLSVFFVIIVAILPIFFYTSNKG 57

Query: 61 EDRAKETSGFYGASSQKVNSSTKKGAS 87

+R ET+GFYGAS+ K KT+KAS

35 Sbjct: 58 SNRQAEITNGFYGASTHK---KTRKAS 80

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1179

- 40 A DNA sequence (GBSx1255) was identified in *S.agalactiae* <SEQ ID 3667> which encodes the amino acid sequence <SEQ ID 3668>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0815(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3669> which encodes the amino acid sequence <SEQ ID 3670>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

-1322-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0107 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 43/64 (67%), Positives = 53/64 (82%)

10 Query: 1 MKVALIPEKCIACGLQQTYSNIPDYQDDGIVKPSDIDNLEKEIPSSDQOTVLAVKSCPTK 60
 MKV++IPEKCIACGLQQTYS++FDY D+GIV FS + + I SD+ +LAVKSCPTK
 Sbjct: 1 MKVSIPEKCIACGLQQTYSLLFDYDNGIVTFSSSSSTSQSICPSDQDAILAVKSCPTK 60

15 Query: 61 ALTI 64
 ALT+
 Sbjct: 61 ALTL 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 **Example 1180**

A DNA sequence (GBSx1256) was identified in *S.agalactiae* <SEQ ID 3671> which encodes the amino acid sequence <SEQ ID 3672>. Analysis of this protein sequence reveals the following:

Possible site: 28

25 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.61 Transmembrane 47 - 63 (41 - 69)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5246 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC36851 GB:L23802 pore-forming peptide [Enterococcus faecalis]
 Identities = 42/130 (32%), Positives = 63/130 (48%), Gaps = 9/130 (6%)

 Query: 7 KIRYHQPELSWAIITYWSIAIAPIFIGLSLLYERTE---IPSQVFLFAIFIVLWIGIFH 63
 K +++WQPEL+ IITYWS +FI L L E I + V V F +F L G
 Sbjct: 3 KQFYWQPELASTIITYWSTFCILFISLILALENNGPYLISNLVMPFPVFAYL---GIA 59

40 Query: 64 RYFVIBEDGYLRIVSNFLRRTKFPFIEDIAKIEVTKSSVTIKFNNHE--RIFYMKRWPK 121
 R F + E L + + R+ P+ I K+ + S+ I + E ++F M+K
 Sbjct: 60 RFSNMTSTS-LIVRDVLMFRKKALPLSQISKVTYNEKSEIFSSSFEGSGKVLMKIKTD 118

45 Query: 122 KYPLDALAIR 131
 FL+AL I+
 Sbjct: 119 SLFLHALKIK 128

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3673> which encodes the amino acid sequence <SEQ ID 3674>. Analysis of this protein sequence reveals the following:

50 Possible site: 28

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.87 Transmembrane 47 - 63 (41 - 69)
 INTEGRAL Likelihood = -3.35 Transmembrane 20 - 36 (18 - 37)

----- Final Results -----

55 bacterial membrane --- Certainty=0.4949 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1323-

The protein has homology with the following sequences in the databases:

```
>GP:AAC36851 GB:L23802 pore-forming peptide [Enterococcus faecalis]
Identities = 42/130 (32%), Positives = 70/130 (53%), Gaps = 12/130 (9%)

5   Query: 7   KIRYHWQPELSWSIIYWSIAFAPIFVGLSLYERTE---IPSRVFILFAIFAVLVGIGLH 63
      K ++WQPEL+ +IIYWS F +F+ L L E I + V + F +FA L G+
      Sbjct: 3   KQKPYWQPELASTIIYWSCTFCILFISLILALENNNGPYLISNLVWVPFVFAYL---GIA 59

Query: 64   RYF-LIENNGILRIVSFKLPGPKLILSTTKIKVTKSLCL---HVEDKSYLFYMRKWP 119
      R F + E + I+R V + F + L +S I K+ + + + + + S +F +K
10  Sbjct: 60   RSEMTETSLIVRDVLW--PRKKALPLSQIKRVTYNGKIRIFSEPKGSGKVFIMKKKT 117

Query: 120   KKYFLDALAV 129
      FL+AL +
15  Sbjct: 118   DGLFLEALKI 127
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 115/162 (70%), Positives = 132/162 (80%), Gaps = 1/162 (0%)

20  Query: 1   MIKLPKIRYHWQPELSWSIIYWSIAFAPITGLSLYERTEIPSCQVFLFAIFVLVGI 60
      MIKLPKIRYHWQPELSW+IIYWSIA APIT+GLSLYERTEIP+VF+LFAIF VLVGI
      Sbjct: 1   MIKLPKIRYHWQPELSWSIIYWSIAFAPITGLSLYERTEIPSRVFILFAIFVLVGI 60

Query: 61   GSRVYFVIEEDGYLRIVSFNPLRRTKFPFIEDIAKIEVTKGSGVTKFNNRHERIFYMRKWP 120
      G HRVFIIE +G LRIVSF K I I KIEVTKS++ + + +FMRKWP
25  Sbjct: 61   GLHRVFIIEENGLIRIVSFKLGPRLILSTTKIEVTKSLCLAVEDK-SYLFYMRKWP 119

Query: 121   KKYFLDALAIEPTFKGEVELLDNLKMDYFRCYRYDKKALTK 162
      KKYFLDALA+ P F+GEV L DN IK+DYFE Y++DKKALTK
30  Sbjct: 120   KKYFLDALAVNPYPQGEVILSDNFIKLEFVYQDKKALTK 161
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1181

35 A DNA sequence (GBSx1257) was identified in *S. agalactiae* <SEQ ID 3675> which encodes the amino acid sequence <SEQ ID 3676>. This protein is predicted to be peptidase t (pepT). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2913 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA20627 GB:L27596 tripeptidase [Lactococcus lactis]
Identities = 274/406 (67%), Positives = 334/406 (81%), Gaps = 4/406 (0%)

50  Query: 1   MSYEKLLERFLIYVKINTRSNPNSTQPTTOSQVDFALTVLKPENRAIGLKDQVHYLSPNG 60
      M YEKL RFL YVK+NTRS+ NST TP+TQ+ V+FA + +H+A+GLKDQVHYL SNG
      Sbjct: 1   MRYEKLPRFLYEVKVNTRSDENSTTTPTQALVEFAHK-MGEDMKALGLKDQVHYLESNG 59

Query: 61   YLVGTLFATSKRLRKIGFISIMDTADFNAINITPQIVDYKGGD--IRLSDSGYILSPKD 118
      Y++GT+PA +D+ KTG ++H+DTADFNAE+ PQI++ G+ I+LSD+ + L PKD
55  Sbjct: 60   YVIGTIPANTDKVRKIGLGLAHLTADFNAGVNPQILENYDGSVYQLGDTFTTIDPKD 119

Query: 119   FPNLNNYHGQTLITTDGKTLGLADKSGIAEINTAMRYLAS-HPEIHCERIVGPGPDER 177
      FPNL NY GQTL+ TDG TLG+LQKSG+AEINT +YL + +P+ EH KIRVGPDPDER
```

-1324-

Sbjct: 120 PPNLKNYKQQTIVHTDGTLLGSDDKSGVARIMTADYLLNLINPDPFGEIRGVGFPDDE 179

Query: 178 IGIGADKDFVDKDFVDFAFYTVDGGPLGELQYETPSAAGLETFEGRNVHPGTAKNMINA 237
IG+GADKFDV DFFVDFAFYTVDGGPLGELQYETPSAAG + P+G+VHFGTAKN M+NA

5 Sbjct: 180 IGVGADKFDVADFFVDFAFYTVDGGPLGELQYETPSAAGVIEPQGVNHVGTAKNMVNA 239

Query: 238 LQLAMPFHSQLEPNERPEQTDGYQGFYHLYDLGCTVDQAKSSYIIRDPEEVDPLKRLHQA 297
LQA+D+H+ LFE +RPE+T+G +GF+HL L GT +A++ YIIRD ER F +RK L

10 Sbjct: 240 LQLAIDYINALPEFDRPEKTEGRGFPFILLKLGTFEKAQAQYIIRDHSGKFWERKALM 299

Query: 298 QDIADNMNEALQSERVKVLYDQYNNMKVKIEKMTPIINIAKEVMEELDIKPIIEPIRG 357
Q+IAD MN L RVK + DQYNNM +LSEKM+ I+RAK+ MS LDI PIIEPIRG

15 Sbjct: 300 QDIADNMNEALQSERVKVLYDQYNNMAQIEKMTPIIDIAKAMENLDIAPIEPIRG 359

Query: 358 TDGSKISFMGIPTNPLFAGCENMHGRFEPVSLQIMEKAVDVLIGIV 403
TDGSKISFMG+PTNPLFAGCENMHGRFEPVSLQIMEKAVD +L I+

15 Sbjct: 360 TDGSKISFMGLPTNPLFAGCENMHGRFEPVSVQIMEKAVDVLILEII 405

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3677> which encodes the amino acid sequence <SEQ ID 3678>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2938 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAs and GBS proteins is shown below.

30 Identities = 305/406 (75%), Positives = 352/406 (86%), Gaps = 1/406 (0%)

Query: 1 MSVEKLLERFLTVVKINTSRNPMSTQTPITTSQVDFALITLKPENAEIGLQDVHYLPNG 60
M Y+ LL+RF+ YVK+NTSR P+S TP+T+SQ FALT+LKPEMAEIGL+DVHY P NG

35 Sbjct: 5 MKYDNLLERFKIKVKNTRSVDPGDTTPTSTESQEAFLITLKPENAEIGLQDVHYLPNG 64

Query: 61 YLAVTLFATSEDLRHKIGFISHMOTADFNAENITPQIVD-YKGGIEILGDSGYLSPKDF 119
YL+GTLFA + L KIGFI+HMTADFNAEN+ PQI+D Y+GGDI LG S Y L PK F

35 Sbjct: 65 YLIGTLFANNPTLTKKIGFIAHMTADFNAENVNPOI IDNYGGGDI TLGSSNYKLDPKAF 124

Query: 120 PNLNMYHGQTLITTDGKTLGADDKSGIARIMTAEMEYLASHPETIEHCIRVGVGFPDEIG 179
PNLNMY GQTLITTDG TLLGADDKSGIARIMTA+E+L S P+IEHC+I+V FGDPDEIG

40 Sbjct: 125 PNLNMYIGQTLITTDGKTLGADDKSGIARIMTAIEFLTSQOQIEHCIDKRVAGFPDEIG 184

Query: 180 IGADKFDVVDKDFVDFAFYTVDGGPLGELQYETPSAAGLETFEGRNVHPGTAKNMINALQ 239
+GADK+V DF+DFAYT+DGGPLGELQYETPSAA LB+TP GRNVHPGTAK+QMINAL+

45 Sbjct: 185 VGADKFEVADFEVDFAFYTVDGGPLGELQYETPSAALBVTPLGRNVHPGTAKDQMINALE 244

Query: 240 LAMPFHSQLEPNERPEQTDGYQGFYHLYDLGCTVDQAKSSYIIRDPEEVDPLKRLHQA 299
LA+DFH +L+ +RFE TDGYQGFYH L+GTV+LA++SYIIRDPEF F RK ++

50 Sbjct: 245 LAIDFHEKLEPAKDRPEYTDGYQGFYHLYGLTGVIEGRASYIIRDPEEASPEARKVKN 304

Query: 300 IADNMNEALQSERVKVLYDQYNNMKVKIEKMTPIINIAKEVMEELDIKPIIEPIRG 359
IA +MN L ++RV V+L DQYNNMKVKIEKMT I +AKRVMEEL IKP+IEPIRG

50 Sbjct: 305 IADNMNEALQSERVKVLYDQYNNMKVKIEKMTPIIEIAKEVMEELDIKPIIEPIRG 364

55 Query: 360 GSKISFMGIPTNPLFAGCENMHGRFEPVSLQIMEKAVDVLIGIVAK 405
GSKISFMGIPTN+L+AGCENMHGRFEPVSLQIME+AVDVI+G+V K

55 Sbjct: 365 GSKISFMGIPTNPLFAGCENMHGRFEPVSLQIMEKAVDVLIGLVCK 410

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1325-

Example 1182

A DNA sequence (GBSx1258) was identified in *S. agalactiae* <SEQ ID 3679> which encodes the amino acid sequence <SEQ ID 3680>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -12.26 Transmembrane 481 - 497 ( 477 - 508)
INTEGRAL Likelihood = -9.45 Transmembrane 510 - 526 ( 505 - 534)
INTEGRAL Likelihood = -7.96 Transmembrane 316 - 332 ( 310 - 334)
INTEGRAL Likelihood = -7.54 Transmembrane 354 - 370 ( 351 - 373)
10 INTEGRAL Likelihood = -7.11 Transmembrane 385 - 401 ( 383 - 409)
INTEGRAL Likelihood = -6.58 Transmembrane 215 - 231 ( 211 - 233)
INTEGRAL Likelihood = -6.48 Transmembrane 71 - 87 ( 69 - 91)
INTEGRAL Likelihood = -6.32 Transmembrane 110 - 126 ( 106 - 133)
INTEGRAL Likelihood = -5.10 Transmembrane 446 - 462 ( 443 - 465)
15 INTEGRAL Likelihood = -3.29 Transmembrane 418 - 434 ( 418 - 435)
INTEGRAL Likelihood = -2.55 Transmembrane 263 - 279 ( 263 - 279)
INTEGRAL Likelihood = -2.02 Transmembrane 142 - 158 ( 141 - 159)
INTEGRAL Likelihood = -1.70 Transmembrane 184 - 200 ( 184 - 200)

20 ----- Final Results -----
      bacterial membrane --- Certainty=0.5904 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

25 A related GBS nucleic acid sequence <SEQ ID 8747> which encodes amino acid sequence <SEQ ID 8748> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crcnd: 6
MoG: Discrim Score: -10.58
GVH: Signal Score (-7.5): -1.1
30 Possible site: 32
>>> Seems to have no N-terminal signal sequence
ALOM program count: 13 value: -12.26 threshold: 0.0
INTEGRAL Likelihood = -12.26 Transmembrane 470 - 486 ( 466 - 497)
INTEGRAL Likelihood = -9.45 Transmembrane 499 - 515 ( 495 - 523)
35 INTEGRAL Likelihood = -7.96 Transmembrane 305 - 321 ( 299 - 323)
INTEGRAL Likelihood = -7.54 Transmembrane 343 - 359 ( 340 - 362)
INTEGRAL Likelihood = -7.11 Transmembrane 374 - 390 ( 372 - 398)
INTEGRAL Likelihood = -6.58 Transmembrane 204 - 220 ( 200 - 222)
40 INTEGRAL Likelihood = -6.48 Transmembrane 60 - 76 ( 58 - 80)
INTEGRAL Likelihood = -6.32 Transmembrane 99 - 115 ( 95 - 122)
INTEGRAL Likelihood = -5.10 Transmembrane 435 - 451 ( 432 - 454)
INTEGRAL Likelihood = -3.29 Transmembrane 407 - 423 ( 407 - 424)
INTEGRAL Likelihood = -2.55 Transmembrane 252 - 268 ( 252 - 268)
INTEGRAL Likelihood = -2.02 Transmembrane 131 - 147 ( 130 - 148)
45 INTEGRAL Likelihood = -1.70 Transmembrane 173 - 189 ( 173 - 189)
PERIPHERAL Likelihood = 1.43 21
modified ALOM score: 2.95

*** Reasoning Step: 3
50 ----- Final Results -----
      bacterial membrane --- Certainty=0.5904 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
55

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA00276 GB:AF008220 YtgP [Bacillus subtilis]
Identities = 178/545 (32%), Positives = 302/545 (54%), Gaps = 26/545 (4%)
60 Query: 24 QNVKGTALITAGNIFGRLLGRIYIPWYAMWGHGAZANALPGMGYRIYALFLIISTVGI 83
+++++GT LT G +ISR+LG +Y+IP+ +G A ALF CY Y LFL I+T+G
Sbjct: 4 KLRGTFTVLTGTIYIRLLGMVYLIPPSIMVG---ATGALFPQYGYNYTLFLNLTATGSP 60

```

-1326-

Query: 84 FVAVAKQVSKYNILGKHEMSIYLVRKILQFMLIGGIFALIMYIGSPILFASLGKGOQ-- 141
 P AV+K VSKYN+ G E S +++ + M+ G I 1+Y+ +P+FA +S GQ++
 Sbjct: 61 PRAVSKFVSKYNKNGDYETSRKMLKAGMSVMLVTGMIAFFILYLSAPFAEISLGCKDNN 120

5 Query: 142 -----LVPIIASLTIAVLVPPMSVLRGFPQGFNNLKFPYAISQVABQIIRVIMMLITAF 195
 +Y ++R ++LA+LV P MS++RGFPQG + P ASQV EQI+R+I++L F
 Sbjct: 121 GLTIDRVVYVIRWMSALLVVPIMSLVRGFPFGHGMGPTAVSQVVEQIVRIIFLSATF 180

10 Query: 196 YIKRLGSGDYIAVQSTPAAPVGMFASIAVLLYFLW--RYNNLSALIGKTPKHILKLDTK 253
 I+++ +G + AV +TFAA +G P + V+LY W R L A++ T L K K
 Sbjct: 181 LILKVFNGGLVIAVGATPAALIGAPGL-VVLYYWKRRKSGSLAMMETGPTAMLSYK 239

Query: 254 EILITETKEAIPFIITGAAIQIFKLIDQPSFGNTM--ALFTNYSSEELRUMPAVFPSSNPG 311
 ++ E A P++ G AI ++ ID +F M A S + L ++ Y
 Sbjct: 240 KMFFELFSTAAPVVFVGLAIPLYNIYDTNTPKQMIETGCHQAI SQDMLAILLYVQ--- 295

15 Query: 312 KVTMILIAVATAIAGVCIPLLTENFVNDKKAARLVNMLQMLIMPLPAVAGSVILAK 371
 K+ M1 +++ATA IP +TE+ F + K + + + +Q +L ++P+V G +L+
 Sbjct: 296 KLVNIPVSLATAPGLTIPTITESTSGNYKLLNQINQTHQITLFLIIPAVGVISLSSG 355

20 Query: 372 PLYTVFYGL----PQQAQLGLFVLSLIQITILSIYTVIAPMLQALFENRKAIIYFLYGLV 427
 P YT FYG --P+ A L S + I+ S+TV A +LQ + + + A++ + G+V
 Sbjct: 356 PTTYTFYGSLSLHPLGAILNLMYSPV-AILPSLPTVNAAILQSIINKQFAVSVLVCVV 414

25 Query: 428 AKVILQLPSILPHAYGFLPSTTVALCIPVIMLYKIHETGPKRARRTSALVILTL 487
 K++L +P I L A G + T + ++ + I G+ + + + + L+L+L+
 Sbjct: 415 IKLVLRVPLIKLMQADGAILATALGYIASLLYGFIMIKRHAGSYKILVIRVIMLVLSA 474

Query: 488 LMSPIISMIMILMLVI-VPD SRLSVLVIIVIGAIIGLVYGFMAATHLLDMIGSKAQ 546
 +M + ++ W++ I D + + + + + A+G VY + L K+G R
 Sbjct: 475 IMGIAKIVQVNLGFPFISYQDGGQRAIVVVVIAAAGVAVLYGVRFLGKILGRRLP 534

Query: 547 DLRRK 551
 RK
 Sbjct: 535 GFPRK 539

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3681> which encodes the amino acid sequence <SEQ ID 3682>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.60	Transmembrane	468 - 484 (466 - 493)
INTEGRAL	Likelihood = -8.39	Transmembrane	305 - 321 (299 - 323)
INTEGRAL	Likelihood = -7.75	Transmembrane	343 - 359 (340 - 362)
INTEGRAL	Likelihood = -6.58	Transmembrane	374 - 390 (373 - 398)
INTEGRAL	Likelihood = -4.25	Transmembrane	138 - 154 (137 - 157)
INTEGRAL	Likelihood = -3.45	Transmembrane	100 - 116 (98 - 122)
INTEGRAL	Likelihood = -3.40	Transmembrane	415 - 431 (410 - 432)
INTEGRAL	Likelihood = -3.35	Transmembrane	499 - 515 (499 - 519)
INTEGRAL	Likelihood = -2.60	Transmembrane	433 - 449 (432 - 451)
INTEGRAL	Likelihood = -2.50	Transmembrane	173 - 189 (173 - 190)
INTEGRAL	Likelihood = -0.59	Transmembrane	201 - 217 (201 - 220)

----- Final Results -----

bacterial membrane ---	Certainty=0.4439(Affirmative) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm ---	Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AMC00276 GB:AF008220 YtgP [Bacillus subtilis]
 Identities = 169/536 (31%), Positives = 295/536 (54%), Gaps = 24/536 (4%)

Query: 14 MVOGAANSTAGNFISRLGLVLYIIPWYIWMQYAIQNALFNMGNVYVYPLLSTTGLN 73
 +++G T G +ISR+LG++Y+IP+ I +G ALF GYN Y FL I+T G
 Sbjct: 5 LLRGTFTVATGTIYSRIIGWVYLIPFSIMVGA---TGQALPQYGVNQYTLFINIATGAFP 61

-1327-

Query:	74	VAAIKQVAKCKNMGCQTEHSYQLIRSTIKMLGLGLIPSAIMYLGSPFASLE-GQDGT--	130
		A++K V+KYN S G E S +++++ + +ML G+I I+YL +P+PA +S G D	
Sbjct:	62	AAVSKFVSCKNSGQDYETSRMLKKGMSVMLVGMIAFFLILSAPMPARISLQCKDNNG	121
5	Query:	131	-----LVPIMHSLGLAVFIPFVMSVIRGIPOCHNNIKPVAVSQIARQLIRVIMMLTTFP
		+V ++ +SLA+ + P+MS++RG PQGH + P AVSQ+ EQ+R+I+L+L TF	185
Sbjct:	122	LTIDHVVYVIRVMSIALLVVPINSLVRSQFQGHQMMGPTAVSQVQVIRLII FLSSATFL	181
10	Query:	186	IMKLGSGDYASAVTQSTPAAFIGVASHGVVLGYIWM--KQLLAAIPSKFDHVTVIDIKG
		I+K+ +G AV +TPRA IG + VL Y W + +G L A+ T ++ K	243
Sbjct:	182	ILKVPFGGLYIAVGATTAALIGAFGLVLVL-YTYNNKRGSLAHMMNTGPTAMLSYKK	240
15	Query:	244	LLLETLKESIPFIVTGSALQAFQLIDQWTFVNTMTLFTDYSRSQ--LVLPGYFNANPAK
		+ E + P+ + G AI + ID TP M + SQ L +L Y K	301
Sbjct:	241	MPFELPSYAAPYVFGIAIPLNYIDITFNFKAMIEAGHQATSQWMLAILLYVQ---K	296
20	Query:	302	ITWVLIAVASIGVGIALLTENTVKKDKKAARLI INNIEMLVMLPALTAIILARP
		+ M+ +++A + G I +TE++ + K + I ++ ++ +PA+ G +L+ P	361
Sbjct:	297	LVMIPVSLATAFGLTLIPTTSTESGNVKKLQQINQMTTLTFLIIPAVGSLSGP	356
25	Query:	362	LYSVTQASE---ERAIHLFVAVLPQTLLALLTLFSPMLQALPENRKAIYYFAYGLIK
		Y+ FYG+ E ++ + +L+L+T+ +LQ + + + A+ G++IK	418
Sbjct:	357	TYTFFGSESLHPELGANTILMYSVALIFSFLTVAAILQGIHQKQVAVSLVGVVTK	416
30	Query:	419	LVLQIPLIYLHAYGFLAATTIALVVPVIMYMYRKNQVTHNRKLLQKRLLTLIETLM
		LVL +PLI L+ A G +LAT + + + + + + K+L ER +L L+ +M	478
Sbjct:	417	LVLNPLIKMADGAILATAIGYIASLLYGFIMIKKHAGYSYKILVGRVIMLVLSTAM	476
35	Query:	479	GLVVFVAMMLLGYAFK+PTGRITSLLYLLIQLGAMVYVYTLTLTHQDLKLGSK
		G+ V + W+LG+ G++ + + +I +G VV L K++G	533
Sbjct:	477	GIAYKIVQWLGFPFISYQDQGMQNAIVVIAAAGVAVILYCGYRGLPKLIGRR	532

An alignment of the GAS and GBS proteins is shown below.

	Identities = 320/541 (59%), Positives = 431/541 (79%)	
35	Query: 12	MSQRTKVSQBBQMVHGTAMLTAGNFISRLGLGAIYII PWYMWGKHAEEANALPGNGYEI 71 MS + ++Q+E M+G AW TAGNFISRLGL +YIIPWY WNG++A +ANALE MGY + Sbjct: 1 MSTEEKQLTQEBLMVQGAANSTAGNFISRLGLVLYII PWYIMWQVATQANALPNNGYIV 60
40	Query: 72	YALFLLISTVGI PVAQVAKQSKYNTLGEKMSIYLVKILQPMILGSGIPALIMYIGSPL 131 YA FLLIST G+ V+AKQV+KYN++G+ E S L+R L+ ML LG IP+ IMY+GSP Sbjct: 61 YAYFLLISTGLVAIAKQVAKTNSMGQTEHSYQLIRSTIKMLGLGLIPSAIMYLGSP 120
	Query: 132	FASLSKGGQSLVFLRSLLTAVLVFSPMSVLRGPFQGFNNLKPVAISQVABQIRVIMWL 191 FASLS G LVPI+ SL+LAV +FP MSV+RG FQG NN+KPTA+SQ+ASQ+IRVIMWL Sbjct: 121 FASLSGGDDTLVPIMHSLGLAVFIPFVMSVIRGIPOCHNNIKPVAVSQIARQLIRVIMWL 180
	Query: 192	LTAPYIMELGSGDYIAAVTQSTPAAFVQMFASIAVLLYPLWRYNMLGALIGKTPKHILKD 251 LT P+IM+LGSGDY +AVTQSTPAAF+GH AS+ VL Y+LW+ +L+AA K + +D Sbjct: 181 LTTFPFIMLGSGDYASAVTQSTPAAFIGVASHGVVLGYIWMKQLLAAIPSKFDHVTVIDIKG 240
50	Query: 252	TKELIETIKELIPFITGAAGIQFKLIDQPSFGNTMALFTNYSSEKIRVMFAPSSNPG 311 K +L+ET+KE+IPFI+TG+AIQ F+LIIDQ+P NIM LFT+YS +L+V+ F YP+NP Sbjct: 241 IKGLLETLKESIPFIVTGSALQAFQLIDQWTFVNTMTLFTDYSRSQWMLAILLYVQ 300
55	Query: 312	KVTMLIAVATAGVGIPLLTENTVKNDEKKAARLVNNILQMLMFLIPAVAGSVILAK 371 K+TM+LIIVA +I GVG LLTEN+VK D KAARL+ +NN+ +ML+MFLLE+ G+ILA+ Sbjct: 301 KITWMLIAVASIGVGIALLTENTVKKDKKAARLI INNIEMLVMLPALTAIILARP 360
60	Query: 372	PLYTVFYGLPQQALGLFVVISLIOTIILSIYTVVLAPMLQALPENRKAIYFYGLVAKVI 431 PLY+VPYG + +A+ LFV L QT+ +L+YT+ +FMLQALPENRKAIY FYG++ K++ Sbjct: 361 PLYSVFYGASEERAIHLFVAVLPQTLLALLTLFSPMLQALPENRKAIYFAYGLIKLV 420
	Query: 432	LQLPSIFLPHANGPLSTFTVALCIPVIMLYLKIHEITGPKRQAIRTSALVILITLMSF 491 LQ+P I+L HAYGFL +T+AL +P+ IMY +++++T F R+ + + L LI TLM Sbjct: 421 LQIPLIYLHAYGFLAATTIALVVPVIMYMYRKNQVTHNRKLLQKRLLTLIETLMGL 480

-1328-

Query: 492 IISMIILWMLNVLIVPDSRLVSLVYIIVIGAILGVYGRMALATHLLDKMIGSRAQDLRRKL 552
 ++ + WL+ P RL SL+Y++IG +G VY + L TH LDK+IGS+A LR+KL
 Sbjct: 481 VVFVWLLGLYAFKPTGRILTSLLYLLIIGLGTVTYALTLTHQLDKLIGSKSLRQLK 541

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1183

A DNA sequence (GBSx1259) was identified in *S.agalactiae* <SEQ ID 3683> which encodes the amino acid sequence <SEQ ID 3684>. Analysis of this protein sequence reveals the following:

10 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.4104 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:BA06290 GB:AP001515 UDP-N-acetylmuramoylalanine-1-D-glutamate-2,
 6-diaminopimelate ligase [Bacillus halodurans]
 Identities = 153/468 (32%), Positives = 237/468 (49%), Gaps = 23/468 (4%)

Query: 33 NVTFNALSDSRQISSDTLFFA-KGATPK-KEVLDSAITAGLSFVSETDYGADIPVLV 90
 N +++ DSR++ LFF KG T +Y A++ G VSR +PV++V
 25 Sbjct: 21 NPDHSHMDSREVVEGGLFFCTIGTITVDGHDYQAQVSGAVVSRPLELSEVPPVVV 60

Query: 91 NDIKQMSLSMSFYNNQNKILKLAFTGKTGTNAYPAYHMLKVNHR-PAMLSMTMT 149
 D ++AM+ ++ FY P N L+L+ TGT GKT + +++ + TM T
 30 Sbjct: 61 RDSRRAMQVATKFKYGEPTINDLQLIGVTGTNGKITITTHLIEKIMQDQGGTGLGTMYTK 140

Query: 150 LDKGSFFKSHLTTPSRLDLFRMATAVENQMTLIMEVSSQAYLTKRVYGLTFDVGVFLM 209
 + G ++ TTPEGL L R A ++ +T +MEVSS A + RV G PDV VF N
 35 Sbjct: 141 I-GHELKETKNTTPESLVLRQTADMKKSGVTAMMEVSSHALQSGRVGCDPFAVAFSN 199

Query: 210 ISPDIHQIEHPTFEDYFFHKLRLME-----NSNAVUVN-----SQMDHFNIVKVEQVEYI 259
 ++PDH+ H T R Y F K LL V+N + D + QV
 40 Sbjct: 200 LTFDHL--YHGTMEYKFAKGLLEAQLNTYQKQVAVINADDPASADFAEMTAQVVTY 257

Query: 260 PHDFYQDY-SENVTSSKAFSPHYVEKLEN-TYDKLIGKFNQENLAAGLACLRGVSI 317
 + D+ +ENV S +F+ E I LIGKF+ N+RA A GV+
 45 Sbjct: 258 GIENEADPQAEVNRITSTGTTFELAAFEERMELSIHLIGKFSVYNVLAAMAAVSGVPL 317

Query: 318 EDIKNGIACIT-VPGRMEVLTYTGAKIIFVDYAHNGDSIKKLLAVVEEHQKGDILIVLGA 376
 ++IK + + V GR E + + VOYAH DSL+ +L V E KGD+ +V+G
 50 Sbjct: 318 QRKKSLEEVKQVAGRPFETVKEDQPPTIVIVYAHTPDSLENVLKTVGRGLAKGDRVRVVG 377

Query: 377 PGKQGSRRKDPGVINDQHFNQLVLTADDNPFDPPLVTSQBSHINRPTVII-DEEE 435
 G+ +R ++ N Q I T+D+P R+P+ I ++ ++I DR+R
 55 Sbjct: 378 GGRDKTKRPVMAELATTAN-QAIPTSNDRSEEPMDILRMEQGAQKGSYLMIEDRKE 436

Query: 436 AIANASTLITCKLDAIILACKGADAYQIKGNRENYSGLLEVAKKYLK 483
 AI A L + D I+IACKG + YQ + ++ D VA++ +K
 Sbjct: 437 AIFKALILAK-RDDIIVIAKGHETYQQPRRTITPD-DRIVAQAIAK 482

- 55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3685> which encodes the amino acid sequence <SEQ ID 3686>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----

-1329-

bacterial cytoplasm --- Certainty=0.4717 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 350/482 (72%), Positives = 399/482 (82%), Gaps = 1/482 (0%)

Query: 1 NITIDKILKILKNDNFREILPHKHYNNWNTQVTFNALSYSRQISSDILPFAKGATPK 60
 10 NITL++L+ILK DENFRE+L + Y+Y++ Q +F LSYDSRQ+ TLFFAKGATPK
 Sbjct: 1 NITIRQLLDILKGDHNFREVLADSGYHYH-QQSFPERLSYDSRQVDQKTLFFAKGATPK 59

Query: 61 KEYLDASITAGLSFVSETDYQADIPVILVNDIKKMSLISMSFYNNPQNKILKLAFTGT 120
 +YL AIT GL Y+SE DY IPV+LV DIKQMSLI+M+FY NPQ KKLKLAFTGT
 15 Sbjct: 60 ADYLKAITNGQLVYISKVDYELGIPVVLVTDIKKMSLIAMAFYQNPQKILKLAFTGT 119

Query: 121 KGKTTAAYFAYHMLGVNHRPANGSLTMTWITLGGKSFYKSHLTPESLDLFRMATAVENQM 180
 KGKTTAAYFAYHMLK +++PAM STMTWITLGGK+FFKS LTPESLDLF MNA V N M
 Sbjct: 120 KGKTTAAYFAYHMLKESYKPAFSTMTWITLGGKTFKKSQTLTPESLDLFAMNASCVTNGM 179

Query: 181 THLIMEVSSQAYLTKRVGLTFTDVGVPFLNISPOHIGPIEHPTFEDYFHKRLMENSRAV 240
 THLIMKVSSQAYL RVVGLTFTDVGVPFLNISPOHIGPIEHPTFEDYFHKRLMENS AV
 20 Sbjct: 180 THLIMEVSSQAYLDRVYGLTFTDVGVPFLNISPOHIGPIEHPTFEDYFHKRLMENSRAV 239

Query: 241 VINSQMDHFNIVKEQVEYIPIHDFYQDYSENVITESKAFSPHVKGKLENTYDIKLIKFNQ 300
 V+NS MDHP+ + +QV H FYG S+N IT S+APSP KG+L YD+GL FNQ
 25 Sbjct: 240 VINSQMDHFSFLADQVADQEHVFFYGLSDNQITTSQAFSFEAKQGLAHYDQLIGHFNQ 299

Query: 301 ENAIAAGLACLRIGVSIHDIKNGIAQTTPVGRMEVLTQNGAKIPVDYAHNGDSLKLLA 360
 ENA+AGLACLRIG S+ DI+ GIA+T VGRMEVLT TN AK+VVDYAHNGDSL+KLL+
 30 Sbjct: 300 ENAIAAGLACLRIGASLADIQKGIATKRVFGRMEVLTWNHAKVFPVDYAHNGDSLKLLS 359

Query: 361 VVEEHQKQDILILVLAGPQNGQSRERKDFGQVINGHPMLQVLTADDFNFDPLVISQRTA 420
 VVEEHQ G ++L+LGAPQNG+SRR DFG VI+QHPL VLTADDFNFDP IS+RTA
 35 Sbjct: 360 VVEEHQTKIMLITAGPQNGKESRADPGRVINGHPMLTVILTADDFNFDPEDISKEITA 419

Query: 421 SHINRPVTIIDREEAIANASTLTNCKLDAILIAGKGADAYQIKNGRDNYSQGLVAKKYL 482
 SHI RPV II DRE+AI A +L DA+IAGKGADAYQI+KG + Y+QGL +AK YL
 Sbjct: 420 SHIARPVETISDRQAQIKAMSLCQGAQDAVITAGKGADAYQIVKQCVAYAGDLAIGHYL 481

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1184

A DNA sequence (GBSx1260) was identified in *S. agalactiae* <SEQ ID 3687> which encodes the amino acid sequence <SEQ ID 3688>. Analysis of this protein sequence reveals the following:

45 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.1421 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1330-

Example 1185

A DNA sequence (GBSx1261) was identified in *S.agalactiae* <SEQ ID 3689> which encodes the amino acid sequence <SEQ ID 3690>. This protein is predicted to be FhuA (fepC). Analysis of this protein sequence reveals the following:

```

5      Possible site: 54
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2785(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9975> which encodes amino acid sequence <SEQ ID 9976> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF98153 GB:AF251216 FhuC [Staphylococcus aureus]
Identities = 141/259 (54%), Positives = 193/259 (74%)

Query: 7  MSHIKAEINIVSYDQKBIINNLSILNQKITTIIGANGCGKSTLLKALTRIHKIKDGTI 66
      M+ + + + + Y IIN L + I + K+T+IIG NGCGKSTLLKAL+R+ +K+G +
20 Sbjct: 1  MNRLHGQVVKIGYGNHTIINKLDVEIPDGKVTSIIGFNGCGKSTLLKALSRLLAVKGEV 60

Query: 67  TIDGHDIAHLPTKEIAKKIALLPQVLEATSGITVYELISYGRFPFHQKYLGNLTMDRSKI 126
      +DG +I TKEIAKKIA+LPQ E +G+TV EL+SYGRFPFHQK G LT +D+ +I
25 Sbjct: 61  FLDGENIHTQSTKEIAKKIALLPQSPVADGLTVGELVSYGRFPFHQKGFRLTAEIKKEI 120

Query: 127  HWAMEMTIVAQFAMRDVDDLGGGGRQKQWIAMALAQDTITFLDEPTTYLDNHSQLEVL 186
      WAME+T F +R ++DLGGGGRQ+VWIAMALAQ DT IFLDEPTTYLD+ HQLS+LE
30 Sbjct: 121 DWAMEVTGDTDFRRHSINDLSGGGRQWVIAMALAQRTDIIIFLDEPTTYLDICHQLEIL 180

Query: 187  LLKKINDETQKFIIMVLMHDLNLSARYSYLVAMKKGKIIYEGSPSQIMTKDIIKIDFKID 246
      L++KLN E TI+MVLHD+N +R+SD+L+AMK G II GS ++T++I++ +F ID
35 Sbjct: 181 LVQKLNQBGCTIVMVLHDINQAIRPSDHLILAMKSGDIIATGSTEDVLTQSEILKRVNID 240

Query: 247  AHHIQDPISKQPVLLSYQL 265
      + +DF + +P+L++Y L
35 Sbjct: 241 VVLSKDPKTKGKPLAVTYDL 259

```

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1929> which encodes the amino acid sequence <SEQ ID 1930>. Analysis of this protein sequence reveals the following:

```

      Possible site: 48
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
45     bacterial cytoplasm --- Certainty=0.2970(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50     Identities = 166/259 (64%), Positives = 208/259 (80%)

Query: 7  MSHIKAEINIVSYDQKBIINNLSILNQKITTIIGANGCGKSTLLKALTRIHKIKDGTI 66
      M+ I AE++ ++Y+Q+ I+ LS I KITTIIGANGCGKRS+LLKALTR+ K G +
55 Sbjct: 1  MTTISAEIDLTTAYBQRTIITDKLSFYIPBGKITTIIGANGCGKSSLKALTRLLLPKQGVV 60

Query: 67  TIDGHDIAHLPTKEIAKKIALLPQVLEATSGITVYELISYGRFPFHQKYLGNLTMDRSKI 126
      ++G +IA L TKE+AKK+ALLPQV EAT GITVYEL+SYGRFPFHQ Y GNL+ D+ I
Sbjct: 61  YLNGQNIATLETKEVAKKIALLPQVQEAINGITVYELVSYGRFPFHQSYPGNLSPADKQAI 120

```


-1331-

Query: 127 HWAMEMTVAQFANRDVDDLSSGGRQKQVWIALAQTDTTIFLDEPTTYLDLNNHQLFVLE 186
 HWAM+ TRV +A++ VD LSGGRQ+VW+AMALAQ TDTTIFLDEPTTYLDL+NHQLF+LE
 Sbjct: 121 HWAMQATNVMAYADQPVDAISSGGRQVWIALAQAQTDTTIFLDEPTTYLDLNNHQLFVLE 180

5 Query: 187 LAKKLNDETORTIIMVLHDLNLSARYSDVLMAMTKGKIYEGSPSQIMTKDIKDIFKID 246
 L+K LN+ KTI+MVLHDLNLSARYSD+L+AMK GK I Y G+ +MT II+DIF+I
 Sbjct: 181 LVKSLNKDAGKTI+MVLHDLNLSARYSDHLIAMGKGKIHNYGTGLADWMSPIIQDIFQIK 240

10 Query: 247 AHIIQDPISKQPVLLSYQL 265
 ++ DPI P+L+YQL
 Sbjct: 241 FVLVDDPHNCFIVLTYQL 259

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1186

A DNA sequence (GBSx1262) was identified in *Sagalactiae* <SEQ ID 3691> which encodes the amino acid sequence <SEQ ID 3692>. This protein is predicted to be ferrichrome ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07609 GB:AP001520 ferrichrome ABC transporter
 (ferrichrome-binding protein) [Bacillus halodurans]
 30 Identities = 94/301 (31%), Positives = 177/301 (58%), Gaps = 11/301 (3%)

Query: 6 IIVLTLLTFFLV---SCQQTQKQBSTKTTISK--MPKIBGFTTYGKIPENPKKVINFTYS 60
 +++LT+L F L+ +CG T E S+ M E T ++P NP++V+
 Sbjct: 7 LLLLTMLEFALLVVAACSSNTDAEQADELESEDEGMITYESTGTGPIEVFANPQVWV--ALG 64

35 Query: 61 YTGVLKGLGVNVSYSLDLEKDSVPVKGKQKLEAKKLTADPTFAIAQKPDLMVFDQPN 120
 +TG +L L VNV K+P + + L++ +++ ++ E I PDLI+ + N
 Sbjct: 65 FTGNLALDVNVVGVD-TWKNKNPVYQLQDVTEVSEENLEQIMELDPDLIIAYSTVQN 123

40 Query: 121 INTLKKIAPTLVIKYGAQNYLDMMFALGKVPKGKEANQWVSWKTKTLVAVKKDLHLHLK 180
 L++IAPT++ Y +YL+ +GK+ KE+EA WV +K + +++ +
 Sbjct: 124 AEQLQIAPTIVLYTNNLDYLEQHIVHIGKILNKKEEAQAWDDFKAFARAEQAGEIRKEIG 183

45 Query: 181 ENITFTTMDPYDKNIYLYGNNGRGGRGLIYDSLGYAPEKVKVKDVFKKGWFTVSQEAIGD 240
 + T +++ ++ +Y++GN+GRG E+Y ++ A PE+V++ G++ +S EA+ +
 Sbjct: 184 EDATVSVIETFDQLYVPGNNGRGRLIYQIMDLMPERVSEALADGYTALSFEALPE 243

Query: 241 YVGDYALVNINKTKKMASSLESQVWKNLPAVKKGHIIESNYDVFPYFSDPLSLAQKLSF 30
 + GDY +++ N +A +S +E++ +++PAV+ G + E+N FYF+DPLSLQ L+ F
 50 Sbjct: 244 FAGDYIILSRN---DEADNSFQETNTYQSIPAVQNGQVFANAKREFYFNDPLSLAQLEFF 301

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3693> which encodes the amino acid sequence <SEQ ID 3694>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1332-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:BA07609 GB:AP001520 ferrichrome ABC transporter
 (ferrichrome-binding protein) [Bacillus halodurans]
 Identities = 112/306 (36%), Positives = 178/306 (57%), Gaps = 3/306 (0%)

10 Query: 2 KKLTLALLTLCITTTITLACNQATNHSNTASKSLSPMPQIAGVTYYGDIPOKPKRVVSLA 61
 K L L L L + + ACG+ +S M T ++P P+RVV+L
 Sbjct: 5 KHLILLTMLLFAALLVVAACGNSITDABQADELSEEDGMITYSETOPIEVNPANPQVVALG 64

Query: 62 STTYGYLKKLDMLNVLGVTSYDKKNPILAKTVKKAQVAATLEAVTTLKPDILVVGSTER 121
 +TG+ ID+N+VGV ++ K NP + ++ +V+ +LE + L PDLI+ ST +
 15 Sbjct: 65 --PTGNILALDVNVVGVDTVSKNNPNYBQLLQOVTEVSRENLEQIMELDOLLITAYSTVQ 122

Query: 122 NIKQLAEIAPVISIEYRKRDYLVLSDFGRIFNKEDKAKWLKOWKTXTAAEYKAVT 181
 N +QL EIAP + Y DYL+ + G++ NKE++A+ W+ D+K + +E+K
 20 Sbjct: 123 NAEQLQEIAPTVLTYYNNLDYLQFVEIGKLAKKEEACQAVDDPKARABQAGEEKEKI 182

Query: 182 GDKATFTIMGLYBKDYLVFGKDWGRGGEIIHQAFHYDAPEKVKTEVFKQGYLSLSQEVLF 241
 G+ AT +++ +E +Y+FG +WGRG EI++Q PE+V+ GY +LS E LF
 Sbjct: 183 GEDATVSIVETPRDQLYVFGNNMGWGRTEILYCTMDLANPERVEEMALADGYALSFEALF 242

25 Query: 242 DYIGDYVVVAEEDDKTGSALYESKLNQSI PAVKGGHVIKVNANVFYPTDPLSLSEYQLBTL 301
 ++ GDY+++ ++D+ ++ E+ +QSIPAV+ V +NA FYF DPLSL QLE
 Sbjct: 243 EFAGDYIIL-SKNDEADNSPCEINTYQSIPAVQNGQVFENAKEPFYNDPLSLQLQLEFF 301

30 Query: 302 REALIS 307
 +E LS
 Sbjct: 302 KEHFLS 307

An alignment of the GAS and GBS proteins is shown below.

Identities = 140/316 (44%), Positives = 212/316 (66%), Gaps = 12/316 (3%)

35 Query: 1 MKKLGIV-LTLLTFLVSCQQTQESTKIT--ISKMPKIEGTYGKIPENPKKVINP 57
 MKK+ +++ L L T L+CG Q S + +S MP+I G TYG IP+ PK+V++
 Sbjct: 1 MKGLTLLTLCITTTITLACNQATNHSNTASKSLSPMPQIAGVTYYGDIPOKPKRVVSL 60

40 Query: 58 TSYTYGILLKLVN---VSSYSLDLEKDSPVFGKQLKEAKLTADDTEATAAQKPDILMV 114
 +YTYGL KL +N V+SY +K +P+ K +K+AK++ A D EA+ KPDIL+V
 Sbjct: 61 ASTYTYGILLKLDMLNVLGVTSY----DKKNPILAKTVKKAQVAATLEAVTTLKPDILV 116

45 Query: 115 FDQDNINLTKKIAPTLLVIKGAQNYLDMMPALGKVGKGEKANQVMSQWCKTKTLAVKDD 174
 + NI L +IAP + I+Y ++Y+L ++ G++F KE +A +W+ WKTKT A +K+
 Sbjct: 117 GSTENIKQLASIAPIVISIEYRKRDYLVLSDFGRIFNKEDKAKGHLKDWKTCTAAEYKE 176

Query: 175 LHRILKPNITFTTFIMDFYDKNITLYGNPNFGRGGLTYDSLYGAPEKVKVDVFKGGHFTVS 234
 + + TPTIM Y+K++YL+G ++GRGE+I+ + Y APEKVK +VFK+G+ ++S
 50 Sbjct: 177 VKAVIGDKATFTDNGLYEKDVYLVFGKDWGRGGEIIHQAFHYDAPEKVKTEVFKQGYLSLS 236

Query: 235 QEAIGDYVDGYALVNINKTKKQASSLKESDVWKNLPAVKKGHIESNYDVYFSDPLSL 294
 OE + DY+GDY +V K S+L ES +W+++PAVKK H+I+ N +VFYF+DPLSL
 55 Sbjct: 237 QEVLDPYIGDYVVAAR--DOKTGSALYRSKLNQSI PAVKGGHVIKVNANVFYPTDPLSL 294

Query: 295 EAQLKSPTKAIKENTIN 310
 E QL++ +AI + N
 Sbjct: 295 EYQLETLREALISEN 310

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1333-

Example 1187

A DNA sequence (GBSx1263) was identified in *S.agalactiae* <SEQ ID 3695> which encodes the amino acid sequence <SEQ ID 3696>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1188

A DNA sequence (GBSx1264) was identified in *S.agalactiae* <SEQ ID 3697> which encodes the amino acid sequence <SEQ ID 3698>. This protein is predicted to be ferrichrome transport permease (permease). Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> May be a lipoprotein

INTEGRAL Likelihood =-12.74 Transmembrane 129 - 145 ( 123 - 150)
INTEGRAL Likelihood =-10.67 Transmembrane 248 - 264 ( 240 - 283)
INTEGRAL Likelihood =-10.14 Transmembrane 205 - 221 ( 196 - 228)
INTEGRAL Likelihood =-5.95 Transmembrane 319 - 335 ( 317 - 336)
INTEGRAL Likelihood =-3.56 Transmembrane 73 - 89 ( 73 - 90)
INTEGRAL Likelihood =-3.19 Transmembrane 288 - 304 ( 288 - 304)
INTEGRAL Likelihood =-2.76 Transmembrane 266 - 282 ( 265 - 283)
INTEGRAL Likelihood =-2.23 Transmembrane 103 - 119 ( 101 - 122)
INTEGRAL Likelihood =-1.01 Transmembrane 158 - 174 ( 158 - 174)

----- Final Results -----
bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF98154 GB:AF251216 FhuB [Staphylococcus aureus]
Identities = 116/313 (37%), Positives = 194/313 (61%), Gaps = 3/313 (0%)

Query: 26 ILFLIGCYASLRFGAINFKTSDELITVLKNPLKNSNAQDVIFDIRLPRIIAAILVGAMSQ 85
++ LI + S G + S + I + N ++ Q++I +IR+PR IAA++VG A++
Sbjct: 28 MILLLTFLISTLIGDKAKIQASTIIKATFNYNPSNCCQNLINLIRIPRNIIAAVIVGALAV 87

Query: 86 AGAIMQGVIRNAIADPGILLINAGGLALVVAAYFLGSMRYSTIILVCLIGSVISCLIVF 145
+GAI+QGVIRN +ADP L+G+N+GA AL + YA L + + ++ LG+++ +V
Sbjct: 88 SGAIIGGVIRNGLADPALIGLNSGSAFALATYAVLFNTSFLILMFAGFLGAILGSAIVL 147

Query: 146 TISYTRQKGYEHLRLILAGAMISTLFTSVGGVVTLYFKLNRTVIGWAGGLSQINWKLI 205
+ +++ G++ +R+ILAGA +S + T++ Q + L P+LN+TV W AGG+S W L
Sbjct: 148 NIGRSRRDGFNPMRIILAGAVSAMLTALSQCIALAFRLNQTIVTWINGVSGTTHSHLK 207

Query: 206 IIAPIIILGLLISQLLAHQITLISLNSVAKALGQRTQLMTAPILLIVFLSASSVALIG 265
P+I + L I ++ QLITL+L ES+AK LQQ ++ L+I + L + +W+ G
Sbjct: 208 WAIPILIGIALFTIITISKQITINLGESLAKGLGQNVIMIRIGLICITIAMLAGIAVATG 267

```

-1334-

Query: 266 TVSFGLIIHPFKLWPKDYRLLLPLIGFSGATFMIVDLSSRIINPSPETISISSIISI 325
 V+P+GL++PH + I DY +LPL G ++ D+ +R + E + +IIS
 Sbjct: 268 QVAFVGLMVPHIARFLGTIDYAKILELTALIGSILVLVADVIARYL---GEAFVGAISF 324
 ..
 5 Query: 326 VGLFCFLWLIRKG 338
 +G+P FL+L++KG
 Sbjct: 325 IGVPFYFLVLVKKG 337

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3699> which encodes the amino acid sequence <SEQ ID 3700>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.09 Transmembrane 256 - 272 (248 - 287)
 INTEGRAL Likelihood = -10.67 Transmembrane 26 - 42 (23 - 46)
 15 INTEGRAL Likelihood = -6.90 Transmembrane 137 - 153 (133 - 157)
 INTEGRAL Likelihood = -5.10 Transmembrane 167 - 183 (166 - 187)
 INTEGRAL Likelihood = -4.57 Transmembrane 213 - 229 (210 - 232)
 INTEGRAL Likelihood = -2.02 Transmembrane 112 - 128 (110 - 131)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.5437 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:AAF98154 GB:AF251216 Phub [Staphylococcus aureus]
 Identities = 99/274 (36%), Positives = 159/274 (57%), Gaps = 1/274 (0%)
 Query: 34 LSPSLCAVYCHLRFGAVALSHQDLNSILFG-KQNGHKANVLLAIPLRPLFGATL/GSAL 92
 LS L + ++ G + + +F + + N++ IR+PR A + G AL
 30 Sbjct: 26 LSMILLTLPISTLIGDAKIQAFTIIEAIFNPNQCCINIEIRIPNIAVFGVAL 85
 Query: 93 AVSGTQMCAITRNPIAEPLGGINAGAGLALVLAAYVPHLHYSLIILLSSLSIAATL 152
 AVSG I+Q +TRN +A+P L+G+N+GA AL L YA +P+ + ++ LG+ L +
 35 Sbjct: 86 AVSGAIQGVTRNGDLPALIGLNSGASFALALTYAVLENTSFLIIMPAGFLGAILGSAI 145
 Query: 153 VPLGSYQSGKGYHQLRLVLGAMVSIILSALQGQITNYHLANAVIGWQAGGLVGVNMQM 212
 V + G++ +R++LGA VS +L+AL QGI + L V W AGG+ G W
 Sbjct: 146 VLMIGRSRRDGFNPMRIILAGAVERNMLTALSQGLAFAFRINQTVTFVTAGGVSVTWSH 205
 40 Query: 213 IGYIAPLIILSLCAQLLSYHIVILSLSESQAKALQKTNLISAVFMILVILSSRAVAI 272
 + + PLI ++L + +S LT+L+L ES AK LGQ +I + +I+ +IL+ AVAI
 Sbjct: 206 LKVAIPLIGIALFIILITKSLQTLINLIGSLAKGLQGNVIMIRGICLIAMILAGIAVAI 265
 45 Query: 273 AGSISFGLVFIPLMGHPPTPHYRYLLPLCAVSG 306
 AG ++P+GL++PH+ + Y +LPL A+ G
 Sbjct: 266 AQVAFVGLMVPHIARFLGTIDYAKILELTALIG 299

An alignment of the GAS and GBS proteins is shown below.

Identities = 158/295 (53%), Positives = 214/295 (71%), Gaps = 1/295 (0%)
 Query: 6 KKLKQKNSNHFWLVPFITLILFLIGCYASIRPGAINFKTSDLITVLKNPLKNSNAQVVI 65
 KK KS+ FWLVF + + Y I RFGA+ DL ++L +N + +V+
 55 Sbjct: 16 KKTQIITKSHIFWLVFVLLSPSLCAVYCHLRFGAVALSHQDLNSILFGK-QNGHKANVLL 74
 Query: 66 PDRLPRIIAAILVGAAMSQAAGINQGVTRNAIDPGLGINAGAGLALVAYAFGLSHV 125
 IRLPR+ A L G+A++ +G IMQ +TRN IA+PGLGINAGAGLALVAYAF+ +H
 Sbjct: 75 LAIRLPLRFGATL/GSALAVSGTQMCAITRNPIAEPLGGINAGAGLALVLAAYVPHLH 134
 60 Query: 126 YSTILFVCLLGSVSLVCLVFTLSYTRKQGYHQLRLVLAGAMVSIILSALQGQITNYHIA 185
 YS I+++ LLGS ++ LVF LSY KGYHQLRL+LAGAM+S L +++GQ +T Y+ L
 Sbjct: 135 YSLIILLSSLSIAATLVFGLSYQSGKGYHQLRLVLAGAMVSIILSALQGQITNYHIA 194
 Query: 186 RTVIGWQAGGLSQINWKMMLIIPIIILGLISQALAHQTLISLNSVAKALQKQTLN 245

-1335-

VIGWQAGGL +NW+N+ IAP+IIL L ++QL+ L+L+ES AKALGQRT L+
 Sbjct: 195 NAVIGWQAGGLGVNWNMGYIAPLILSLCLAQLSYHLTVLSLSESQAALGQRTNLI 254

Query: 246 TAFLLILVLFLSASSVALIGTVSFIGLIIPHFIKLFIPKDYRLLLPLIGFSGATF 300
 +A +++VL LG++VA+ G+SFIGL+IPH +K F P YR LLEPL SGA+ F
 Sbjct: 255 SAVFMILVLFLSSAAVAIAGSISFIGLVIPHLMKHFTPEHYRYLLPLCAVSGASF 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1189

A DNA sequence (GBSx1265) was identified in *S.agalactiae* <SEQ ID 3701> which encodes the amino acid sequence <SEQ ID 3702>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1492 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 1190

A DNA sequence (GBSx1266) was identified in *S.agalactiae* <SEQ ID 3703> which encodes the amino acid sequence <SEQ ID 3704>. This protein is predicted to be ferrichrome transport permease (permease). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -10.35	Transmembrane	282 - 298 (279 - 309)
INTEGRAL	Likelihood = -7.06	Transmembrane	120 - 136 (115 - 141)
INTEGRAL	Likelihood = -7.01	Transmembrane	62 - 78 (61 - 80)
INTEGRAL	Likelihood = -6.10	Transmembrane	250 - 266 (241 - 272)
INTEGRAL	Likelihood = -5.52	Transmembrane	196 - 212 (190 - 215)
INTEGRAL	Likelihood = -5.47	Transmembrane	155 - 171 (151 - 174)
INTEGRAL	Likelihood = -4.99	Transmembrane	304 - 320 (303 - 322)
INTEGRAL	Likelihood = -3.35	Transmembrane	91 - 107 (90 - 110)

----- Final Results -----
 bacterial membrane --- Certainty=0.5140 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98155 GB:AF251216 FhuG [Staphylococcus aureus]
 Identities = 122/334 (36%), Positives = 208/334 (61%), Gaps = 3/334 (0%)

Query: 1 MIQKNKAPFVLISVVILLILLV---SISLGYANTSVIDVLKLSGKSDDAFLFIITNI 57
 MI N LI+ + +LL L SI+ G N V K + G + D I+ +
 Sbjct: 1 MISSNNGRQLIALAVPSILLPLGCTWSITSGEYNIPIVERFFKTLIGGQDAIDELLDLP 60

Query: 58 RLPRIIVICOGASLGIAGLLQLTATKNPLADSGILGINAGLVIALTIGTFNVSNFTI 117
 RLPR+++ I GA+L I+G +Q++TKNP+A+ GILGINAG G IAL I ++

-1336-

Sbjct: 61 RLPRMHTILAGAALSISGAIVQSVTKNPIAEPGILGINAGGGFAIALFIAIGKINADNF 120

Query: 118 LYFLPLFAMFGGLVTFILYLYMSYRNNHISPTRLIVTGIGISTISGVWILLIISQNNQ 177
+Y LFL ++ GG+ T +I++ S+ +W ++P +++ G+G+ T + G I I+S+ +++

5 Sbjct: 121 VVVVPLISILGIGTIALIIFIPSPNNKRGVTPASWVLIGVGLQTALYGGSSITIMKFTDK 180

Query: 178 KMDMIVEMLSGKITISSWTTIITFIPILILKGLAYSRSRHLNIMNLNEQTALALGLHLK 237
+ D I W +G I W +I F+P +++ +S LNI++ + A LG+ L

10 Sbjct: 181 QSDFLAANFAGNIGWDEMPFVIAFLPWLVIITIPYLPKSNLTNIIHTGDNIARGVLRS 240

Query: 238 KERIYTLMTSSLAASVVLIGNITFIFGLLAGHLRRLLGNHKKIILPSCLLIGAILLV 297
+ER+ + L++ +V + G+I+FIGL+ H++R++G H++ LP +L+GA +I++

10 Sbjct: 241 RERLILFFIAVMLSSAAVAVAGISIFIGLMSPHIAKRIVGPRHQLFLPIALVGLACLVI 300

15 Query: 298 SDTIGRLLLVGTGIPNGLVSVIIGAPYFLVMK 331
+DTIG+++L G+P G+V+IIGAPYFL+LM K

Sbjct: 301 ADTIGIVLQPGGVPAIVVLIIGAPYFLVMK 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1939> which encodes the amino acid sequence <SEQ ID 1940>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.93	Transmembrane	254 - 270 (252 - 284)
INTEGRAL	Likelihood = -10.46	Transmembrane	294 - 310 (292 - 320)
INTEGRAL	Likelihood = -6.74	Transmembrane	25 - 41 (18 - 43)
INTEGRAL	Likelihood = -6.26	Transmembrane	103 - 119 (102 - 125)
INTEGRAL	Likelihood = -3.66	Transmembrane	164 - 180 (164 - 186)
INTEGRAL	Likelihood = -3.03	Transmembrane	209 - 225 (207 - 226)
INTEGRAL	Likelihood = -2.71	Transmembrane	74 - 90 (74 - 91)
INTEGRAL	Likelihood = -2.13	Transmembrane	326 - 342 (325 - 343)
INTEGRAL	Likelihood = -1.97	Transmembrane	135 - 151 (135 - 151)

----- Final Results -----

bacterial membrane	--- Certainty=0.5373 (Affirmative) < succ>
bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/322 (47%), Positives = 229/322 (70%), Gaps = 1/322 (0%)

40 Query: 11 LISSVILLILLIL-VSILGYANTSVIVDLKLSGKSDDAFLPIITNIRLPRIVCIPG 69
L +S+I+L+ ++ +++SLG ++ S +D++ + GKS A FI+ NIRLPRI+ GG

Sbjct: 22 LYTSILLILVSLMLALSGLSHLSPLDLVHVFLGKSSHAISFTIVINIRLPRIACLGG 81

45 Query: 70 ASIGIAGLLIQTITNPLADSGILGACNGLVIALTIGTFNVNSNTIYFLPLFAMFG 129
SL ++GLLIQ LT+NPLADSG+LGI GAG+ +A+ + I +L+PLFAM G

Sbjct: 62 GSLASGLLLQLTRNPLADSGVLGITIGAGISLAIWVSFPFQAHISHTYFLPAMIGA 141

50 Query: 130 LVTFILYLYMSYRNNHISPTRLIVTGIGISTISGVWILLIISQNNQKMDMIVEMLSGK 189
+VT P +Y +S + I FTRLI+TQ+ ++T++S +M+ ++ N K+D+++ VLSG+

Sbjct: 142 IVTFISVYWLSLTRQQQIDPRLILGVAVTMLSSLMVNLVGHINRYKVDVLNWLGGQ 201

55 Query: 190 ITISSWTTIITFIPILILKGLAYSRSRHLNIMNLNEQTALALGLHKKERIYTLMTSS 249
+ M T+ P+L+ W L YS+ + LNI M L+ T+ LGL L +R L+L +

Sbjct: 202 LIGDWTPLSVIAPLLCFLITTSQAHFLINMGADNTAIGLGLPIAKRRRLIIVLAAG 261

Query: 250 LAASVVLIGNITFIFGLLAGHLRRLLGNHKKIILPSCLLIGAILLVSDTIGRLIATV 309
L A+S+I+GNI+FIGL+AGH S L+G+NHKI +P +LIG I+LIV+DT+GR+ LVG+

Sbjct: 262 LGALSVLVGNISFIFGLAGHFSYTLNGSNHKKITIFISILIGMILLVADTVGRVIVLGS 321

60 Query: 310 GIPTILVSVIIGAPYFLVMK 331
I TG++V+IIGAPYFL+LM K

Sbjct: 322 NIQTGILVSLIIGAPYFLVMK 343

65 There is also homology to SEQ ID 1936.

-1337-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1191

A DNA sequence (GBSx1267) was identified in *S. agalactiae* <SEQ ID 3705> which encodes the amino acid sequence <SEQ ID 3706>. Analysis of this protein sequence reveals the following:

Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.3785 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA05779 GB:AF051356 unknown [Streptococcus mutans]
Identities = 49/93 (52%), Positives = 63/93 (67%)

Query: 1 MILTFNPGKLERQEFFKELINYLMIHDDVTLRKIKSHPTDYSKIDRLLEEVINHGYYLRQ 60
MI +N KL RQ FF +LINV L IHDDVTLR+IK +F D ++R +E+Y+ GY+LR+
Sbjct: 1 MIKTYNGDKLTRQFFFIKELINYLQIHDDVTLRQIKRNFALTEHLERSIEDYVQAGYVLR 60

Query: 61 NKRYSLNLPFLSSLDGLVLDLVPIDSDSQIYQ 93
NK Y L +LDGL LD +F+D S IYQ
Sbjct: 61 NKHYNAFELLENLGLTLDLSDQIFVDDQSSIYQ 93

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3707> which encodes the amino acid sequence <SEQ ID 3708>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.3447 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 108/212 (50%), Positives = 143/212 (66%)

Query: 1 MILTFNPGKLERQEFFKELINYLMIHDDVTLRKIKSHPTDYSKIDRLLEEVINHGYYLRQ 60
MI F+ KL RQ FF+LINV L HD V LR+IK F + + ID+ +E Y+ GYI R+
Sbjct: 1 MITVFNPGKLERQEFFKELINYLQIHDDVTLRRIKAFPNVTDGIDKAIESTYVQAGYIRRE 60

Query: 61 NKRYSLNLPFLSSLDGLVLDLVPIDSDSQIYQLQKRFVTHLEINPTMHLVFPVETDPE 120
NKRY +NLP +SS L LD ++F+D+ S +Y+ + F T L N TH ++ E+T+
Sbjct: 61 NKRYGSLNLPVSSDQLALVTLFVDTCSAMYENILAVVFETQLINQTRVRMIKEKINIT 120

Query: 121 RNFITLGNVPYKLTNGYPLSRQKLNQLIGDVNSRYALKOMSSFILKPLKRDQSVQKRT 180
R+ LTL+NYPY+L G S EQ LY LLGVN RYALKOM++F+LKF RKD V QKR
Sbjct: 121 RDDLTLNYPYPLSRGKPSAEQMDLYLLGVNRYALKOMITFLKPKRDKFVQKRP 180

Query: 181 VPIQALELLGYSINQDQTTVRIANGLVEAL 212
IP++AL LGY+ + TTY+L LD E+L
Sbjct: 181 DIPVEALVTLGSLKQVEPTTYQLMITLEKESL 212

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1338-

Example 1192

A DNA sequence (GBSx1268) was identified in *S.agalactiae* <SEQ ID 3709> which encodes the amino acid sequence <SEQ ID 3710>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0824 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAE39104 GB:U57759 intragenetic coaggregation-relevant adhesin
[Streptococcus gordonii]
Identities = 261/311 (83%), Positives = 283/311 (90%)

Query: 1 MSKILVFGHQNPDSDAIGSSVAFAYLAKEAMGLDTEAVALGTPNKEETAYVLDYFGVQAPR 60
      MSKILVFGHQNPDSDAIGSS AFAYLA+EA+GLDTEAVALG PNEETA+VLDYFGV APR
Sbjct: 1 MSKILVFGHQNPDSDAIGSSVAFAYLAREAYGLDTEAVALGEPNEETAFLVDYFGVQAPR 60

Query: 61 VVESAKAGQVETVILTDINEFQCSISDIKDVTYGVVDHHRVANFETANPLYMRLESPVGS 120
      V+ SAKAG E VILTDINEFQCS++DI +V VYGVVDHHRVANFETANPLYMRLESPVGS
Sbjct: 61 VITSAKAGAEQVILTDINEFQCSVADIAEVEVYGVVDHHRVANFETANPLYMRLESPVGS 120

Query: 121 ASSIVYRMFKENGVSVPKELAGLLSGLISDTLLKSPPTTHASDIPVAKELAEAGVNLE 180
      ASSIVYRMFKR+ V+V K3+AG+LSGLISDTLLKSPPTH +D +A ELRLAGVNLE
Sbjct: 121 ASSIVYRMFKESVAVSKELAGLLSGLISDTLLKSPPTHTPKAELAEAGVNLE 180

Query: 181 EYGLMLKAGTNLSKXTAAELIDIDAKTFELNGEAVRVQVNTVDINDILARQEBIEVAI 240
      EYGL MLKAGTNL+SK+A ELIDIDAKTFELNG VRVQVNTVDI ++L RQ BIE AI
Sbjct: 181 EYGLAMLKAGTNLASKSABELIDIDAKTFELNGNNVRVQVNTVDIAEVLERQEBIEVAI 240

Query: 241 QEAIVTGYSDFVLMITDIIVNSSEILAIGSNMAKVEAAF EPTLENNHAFLAGAVSRKQK 300
      ++AI GYSDFVLMITDI+NNSEILA+GSNM KVEAAF F LENNHAFLAGAVSRKQK
Sbjct: 241 EKAIADNGYSDFVLMITDIIVNSSEILAIGSNMDKVEAAFNFVLENNHAFLAGAVSRKQK 300

Query: 301 VVPQLTSYNA 311
      VVPQLTS+NA
Sbjct: 301 VVPQLTSFNA 311

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3711> which encodes the amino acid sequence <SEQ ID 3712>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.02 Transmembrane 141 - 157 ( 141 - 157)

----- Final Results -----
      bacterial membrane --- Certainty=0.1808 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9103> which encodes the amino acid sequence <SEQ ID 9104>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.02 Transmembrane 139 - 155 ( 139 - 155)

----- Final Results -----
      bacterial membrane --- Certainty= 0.181 (Affirmative) < succ>
      bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

```


An alignment of the GAS and GBS proteins is shown below.

Identities = 253/311 (81%), Positives = 283/311 (90%)	
5	Query: 1 MSKILVFGHQNPDSDAIGSSVAFAYLAKBAMGLDYKRAVALGTNNKETAIVLDYFGVQAPR 60 MSKILVFGHQNPDAI SS AF YL+++A+GLDTE VALGTENBETA+ LDYFGV+APR Sbjct: 3 MSKILVFGHQNPDTLALASSTAFDYLKQKAPGLDTEVALGTENBETAFAIDYFGVEPR 62
10	Query: 61 VVESAKAEGVETVILTDHNEFQGSISIDIKDVTYGVVDHHRVANFETANPLYMRLEPVGS 120 VVESAKA+G E VILTDHNEFQGS+DI++V YYGVDHHRVANFETANPLYMR+BPVGS Sbjct: 63 VVESAKAQSEQVILTDHNEFQGS+DIADIEVEVYGVVDHHRVANFETANPLYMRVSPVGS 122
15	Query: 121 ASSIVTRMFKENGVSVPKELAGLLSGLISDTLLKSPFTHASDIPVAKELAEAGVNL 180 ASSIVTRMFKENG+ VPK +AG+LLSGLISDTLLKSPFTH SD VA+ELAEI A VNL Sbjct: 123 ASSIVTRMFKENGIEVPKADGMLLSGLISDTLLKSPFTHSDHVAEELAEAEVNL 182
	Query: 181 EYGLEMLKAGTNLSSKTAELIDIDAKTFELNGEAVRVAQNTVDINDILARQSEIEVAI 240 +YG+ +LKAGTNL+SK+ ELI IDAKTFELNG AVRVQNTVDI ++L RQE IE AI Sbjct: 183 DYGMALLKAGTNLASKSEVELIGIDAKTFELNGAVRVQNTVDIAEVLERQEAIEAI 242
20	Query: 241 QRAIVTEGYSDFVLMITDIVNSNREILA+G+NM KVEAAF PTL+NNHAFAGAVSRKKQ 300 ++A+ EGYSDFVLMITDIVNSNREILA+G+NM KVEAAF PTL+NNHAFAGAVSRKKQ Sbjct: 243 KDAFAEGYSDFVLMITDIVNSNREILA+G+NM KVEAAF PTL+NNHAFAGAVSRKKQ 302
25	Query: 301 VVPQLTESYDA 311 VVPQLTES+ A Sbjct: 303 VVPQLTESFGA 313

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1193

A DNA sequence (GBSx1269) was identified in *S. agalactiae* <SEQ ID 3713> which encodes the amino acid sequence <SEQ ID 3714>. Analysis of this protein sequence reveals the following:

Possible site: 20	
35	>>> Seems to have no N-terminal signal sequence
----- Final Results -----	
	bacterial cytoplasm --- Certainty=0.2769 (Affirmative) < succ>
	bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40	bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05773 GB:AF051356 pyruvate-formate lyase activating enzyme [Streptococcus mutans]	
Identities = 184/260 (70%), Positives = 217/260 (82%)	
45	Query: 3 EIDYKGVKTMHSTSSKGSVDGGRIRFLIFMQGCKMRQCYNHPTWEMETNNSKETVE 62 ++DY+KVTG++STESFGSDVGGRIF++FMQGC+MRQCYNHPTW N+ + + ERT Sbjct: 4 KVDYKGVKGLVNSTESFGSDVGGRIRFLVFMQGCCKMRQCYNHPTWAMKHIDRATERTAG 63
50	Query: 63 DVLKEALRYKHPWKGDDGITYSGGEAMLQIDPITALPIRAKKLGIHITLDYCGFAYRATP 122 DV KEALR+K FWG GGITYSGGEA LQ+DF+ ALP AK+ GLHITLDTC +R TP Sbjct: 64 DVFKEALRFKDFWGTGGITYSGGEATLQNDLIALPSLAKEKGIHITLDYCALTFRTP 123
55	Query: 123 EYHAILKEILLVDLVLVDLKEIDSEKQIKVTRQSNKTLQFARYLSDRGTFPWLIRVLV 182 +Y EKL+ VIDLVLVD+KEI+ +QKIKVT SNK IL ARYLSG Q FWPILIRVLV Sbjct: 124 KYLEKYELKMAVDLVLVDLKEINPDQKIKVTGSEKNTLILACARYLSDIGKFWKIRVLV 183
60	Query: 183 PGLTIDIDLKRIKRGFVCTLDNVDKFEVLVPIYTWGEPKWRRLGIPYPLAGVKPPTPEPVK 242 PGLTD D+ L +LGE+V+TL NV +FE+LPIYTWGEPKWRRLGIPYPL GVKPPT+EV+ Sbjct: 184 PGLTDREDDLKIGETVKTLDNVQRFELPIYTWGEPKWRRLGIPYPLGSKVKEPTDQVR 243

-1340-

Query: 243 NAKDIMKTESYTEYLKRIQN 262
 NAK +M TE+Y EY KRI +
 Sbjct: 244 NAKKLMHTESYTEYKRIINH 263

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3715> which encodes the amino acid sequence <SEQ ID 3716>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4614 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 223/260 (85%), Positives = 239/260 (91%)

Query: 1 NAEIDYKKVIGMIHSTESFGSDGPGIRFIIFMGCGHMRQCYCHNPOTWENETNSKERT 60
 M E DY +VIGM+HSTESFGSDGPGIRFIIF+CGCK+RCQYCHNPOTWENETNSK RT
 Sbjct: 25 NTEKDYGGVGMVHSTESFGSDGPGIRFIIFLQCGCKLRQCYCHNPOTWENETNSKIRT 84
 Query: 61 VDVKLKALRYKHFPGKGGITVSGGEAMLQIDFITALPTEAKKLGINTTLTGTGFAIRA 120
 Y DVKLKAL+YKHFPGK GGITVSGGEAMLQIDFITALPTEAKKLGINTTLTGTG YR
 Sbjct: 85 VNDVLKALQYKHFPGKGGITVSGGEAMLQIDFITALPTEAKKLGINTTLTGTGTYRP 144
 Query: 121 TPEYHAILLEKLLDVTLVLLDLKEIDSEQHKIVTRQSNKILQFARYLSDRGTPWIRHV 180
 TPEYH +L+ LL VTL+LLDLKEID +QHKIVTRQ SNKILQFARYLSD+ P WIRHV
 Sbjct: 145 TPEYHQLVNLAVTLVLLDLKEIDEKQHKIVTRQPNILQFARYLSEKQIFVWIRHV 204
 Query: 181 LVPGLTIDIDHLKRLGSEFVQTLNVDKFEVLYHTNGSEFKWRELGIPIYPLAGVKPPTPR 240
 LVPGLTIDIDHL RLGEFV+TL NVDKFEVLYHTNGSEFKWRELGIPIY L GVKPT ER
 Sbjct: 205 LVPGLTIDIDHLRLGSEFVATLQNVDKFEVLYHTNGSEFKWRELGIPIYQLEGVVKPTPR 264
 Query: 241 VQAKDIMKTESYTEYLKRI 260
 V+NAK+M+TESYTEY+ RI
 Sbjct: 265 VQAKNLMQTESYTEYKRI 284

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1194

A DNA sequence (GBSx1270) was identified in *S.agalactiae* <SEQ ID 3717> which encodes the amino acid sequence <SEQ ID 3718>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -7.06 Transmembrane 105 - 121 (103 - 126)
 INTEGRAL Likelihood = -5.57 Transmembrane 137 - 153 (136 - 162)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3824 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA05772 GB:AF051356 putative hemolysin [Streptococcus mutans]
 Identities = 347/445 (77%), Positives = 406/445 (90%), Gaps = 1/445 (0%)
 Query: 1 MQDPGQSLLQFVILLITLFPNAPPSAGMALVSLNRKSVQKAREGDKRYRRLLDVLE 60
 M+DPGQSLL+LQF++LLILITL NAFPSA+EMALVSLNR++VEQKAREG+K+Y RLL VLS

-1341-

Sbjct: 1 MEDPQSQSLIQFLILLITLCAFFSATMALVSLNRRVQKAREGKKYIRLLKVL 60

Query: 61 NPNFPLSTIQGITIFISLLSGASISASLGKVISGWLGN SATARTAGSIIALFLTYVSIV 120
NPNFPLSTIQGIT I+LL GASL+ SLG I+ W GNSATARTAGS+I+L FLTY+SI+V

5 Sbjct: 61 NPNFPLSTIQGITLITLISGASLADSLGREIAVWFGNSATARTAGSIIASLAFITYSIV 120

Query: 121 LGELYPKRIAMNLDKRIAVSAPIIIFLGKIVSPFWLLSASTNLLSRITPMTPDDADEK 180
LGELYPKRIAMNLIK+ IA++SAP+IIIFLGK+VSPFWLLS STNLLSR+TPMTDDADEK

10 Sbjct: 121 LGELYPKRIAMNLDKRIAVSAPIIIFLGKIVSPFWLLSASTNLLSRITPMTPDDADEK 180

Query: 181 MTRDEIERYMLTNSSEITLDAEIRMLQGVPSLDLMAREVMVPRTDAPMIDINDDSSDIQ 240
MTRDEIERYMLTNSSEITL+AIEMQLQGVPSLDLMAREVMVPRTDAPM+DIN+D+ I+
Sbjct: 181 MTRDEIERYMLTNSSEITLDAEIRMLQGVPSLDLMAREVMVPRTDAPMIDINDDSSDIQ 240

15 Query: 241 GILSQNPSRVFVDDDKDRVGVIVITKRLLEAGFKTGPFITIDRLKIQELFVETIPVD 300
IL++ PSR+PV+DDDKD++G++HTK LL AQPK GFD I+LR+ILQELFVETI V+
Sbjct: 241 TILNERPSRIPVYDDDKDKIIGIHTKRLLEAGFKGFDHILRRILQELFVETIVVN 300

20 Query: 301 DLLKALRNQCNQMAILDEYGVAGLVITLEDLLEIVGEIDDETQARQVREIDENIYI 360
DLL AL+NRQCNQMAILDEYGVAGLVITLEDLLEIVGEIDDETQ VREI +N YI
Sbjct: 301 DLLTALRNQCNQMAILDEYGVAGLVITLEDLLEIVGEIDDETQKTAISVREIADNTYI 360

25 Query: 361 VLGTMTLNDPNEYFETDLSNDVVTIAGYILGVGSIPIQREKVAVEVSDSKKHITLND 420
VLGTMTLNA+PN+YFET+LSD+VVTIAG+YLTGVG+IP+QEEK +EV+S KH+ LIND
Sbjct: 361 VLGTMTLNDPNEYFETDLSNDVVTIAGYILGVGSIPIQREKHEFVSEBNGKLELIND 420

Query: 421 KVKDGRITKLVKLLSDIQ-NIEKD 444
KVKDGR+TKLK+L+S++E+ RKD
Sbjct: 421 KVKDGRVTKLVKILVSEVEKEDKD 445

30

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3719> which encodes the amino acid sequence <SEQ ID 3720>. Analysis of this protein sequence reveals the following:

Possible site: 42

35 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.76	Transmembrane	22 - 38 (16 - 47)
INTEGRAL	Likelihood = -5.57	Transmembrane	118 - 134 (117 - 138)
INTEGRAL	Likelihood = -3.19	Transmembrane	150 - 166 (149 - 169)

40 ----- Final Results -----

bacterial membrane	---	Certainty=0.4503 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

45 The protein has homology with the following sequences in the databases:

>GP:AAC05772 GB:AF051356 putative hemolysin [Streptococcus mutans]
Identities = 343/443 (77%), Positives = 401/443 (90%)

Query: 14 MEDPVSQSLIQFLILLVLELLNAFFSASEMALVSLNRRVQKADGKKYARLLKVL 73
MEDP QSIL+QFLLL+LTL NAFFSA+EMALVSLNR+RVQKLA +G+KKY RLL+VLK

50 Sbjct: 1 MEDPQSQSLIQFLILLITLCAFFSATMALVSLNRRVQKAREGKKYIRLLKVL 60

Query: 74 EPNHFLSTIQGITIFISLLSGASISASLGKVISGWLGN SATARTAGTIIISLFLTYVSIV 133
PN+FLSTIQGIT I+LLSGASL+ SLG+ I+ W GNSATARTAG+ISL FLTY+SI+V

55 Sbjct: 61 NPNFPLSTIQGITLITLISGASLADSLGREIAVWFGNSATARTAGSIIASLAFITYSIV 120

Query: 134 LGELYPKRIAMNLDKRIAVSAPIIIFLGKIVSPFWLLSASTNLLSRITPMTPDDADEK 193
LGELYPKRIAMNLIK+ IA++SAP+II LG+VSPFWLLS STNLLSRITPMTPDDADE+

60 Sbjct: 121 LGELYPKRIAMNLDKRIAVSAPIIIFLGKIVSPFWLLSASTNLLSRITPMTPDDADEK 180

Query: 194 MTRDEIERYMLSKSEATLDAREIRMLQGVPSLDLMAREVMVPRTDAPMIDINDDPLENIQ 253
MTRDEIERYML+ SE TLDA+IEMQLQGVPSLDLMAREVMVPRTDAPM+DINDD +IQ

Sbjct: 181 MTRDEIERYMLTNSSEITLDAEIRMLQGVPSLDLMAREVMVPRTDAPMIDINDDSSDIQ 240

65 Query: 254 EILKQSPSRIPVYDDDKDKIIGIHTKRLLESGFROGFDINRMRLQELFVETIPVD 313

-1342-

```

      IL + PSRIPVYD DKKIIG+IHTK LL +GF++GPD IN+R++LQELFVFPETI V+
Sbjct: 241 TILNRPFRIPVYDDDKKIIGI IHTKNLNLNAGPKRGPFHINLRILQELFVFPETIVVN 300

5  Query: 314 DLLRQLRNTQNMQMAILDEYGGVAGLVLTEDLLEIVGEIDDETQKAEQVHIEGNTYI 373
      DLL L+NTQNMQMAILDEYGGVAGLVLTEDLLEIVGEIDDETQK V EI DNTYI
Sbjct: 301 DLLTALRNTQNMQMAILDEYGGVAGLVLTEDLLEIVGEIDDETQKTAISVREIADNTYI 360

10 Query: 374 VVGTMFLNEFNDFYFTELESDDVDVTIAGFYLTGIGTIPQEQKEAYEIDNKKHVLIND 433
      V+GTMFLN+FN+YF+T+LESD+VDTIAGFYLTG+GTIPQEQ+KE +E+++ KHL LIND
Sbjct: 361 VLGTMFLNDNFYFETDESDDVDVTIAGFYLTGVTGTPQEQKEHFEVESNKKHLELIND 420

      Query: 434 KVKDGRITKLLKLSNIRQIIEE 456
      KVKDGR+TKLK++S +E+ +E
Sbjct: 421 KVKDGRVTKLKILVSEVSEKEDE 443

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 364/444 (81%), Positives = 417/444 (92%)

20 Query: 1 MQDPGQSLLQFVILLILTLNAPFSASEMALVSLNRSKVEQKAEQDKKYKRLDLVLE 60
      M+DP SQSL++QF++L++LTL NAFPSASEMALVSLNRS+VEQKA +GDK+Y RLL VLE
Sbjct: 14 MEDPVSQSLVIQFLLVVLTLNAPFSASEMALVSLNRSRVQKAAADGDKYKARLRLVLE 73

      Query: 61 NPNFLSTIQVGITFISLLQASLASISLGHVIGWLGNSATARTAGSIIALIFLYTVSV 120
      PN+FLSTIQVGITFISLL GASLASLG VISGWLGN SATARTAG+II+L+FLTYVSV
25 Sbjct: 74 EPNHFLSTIQVGITFISLLSGASLASLGKVISGWLGN SATARTAGTII+SLVFLTYVSV 133

      Query: 121 LGELYPKRIAMNLDKRLAIVSAPIIIFLGKIVSPFVWLLSASTNLLSRITPMTFDDADEK 180
      LGELYPKRIAMNLDK+LAIVSAPIII LG++VSPFVWLLSASTNLLSR+TMTFDDADE+
30 Sbjct: 134 LGELYPKRIAMNLDKLAIVSAPIIIFLGRLVSPFVWLLSASTNLLSRITPMTFDDADEQ 193

      Query: 181 MTRDEIEYMLTINSEETLEAEIEMLQGISLDEMMAREVMVPRTDAPMIDINDDAQSNIE 240
      MTRDEIEYML+ SE TL+AEIEIEMLG+FSLDEMMAREVMVPRTDAPMIDIN+D NI+
Sbjct: 194 MTRDEIEYMLSKSEATLDAEIEIEMLGVSFLDEMMAREVMVPRTDAPMIDINDDPLENIQ 253

35 Query: 241 GILSQNFSRPVPFDDDKDRVVGVLHTYKLLLEAGFKTGPDITDLRKILQELFVFPETIFVD 300
      IL Q+PSR+PV+D DKD+++G++HTRKLL+GF+ GPD I++RK+LQELFVFPETIFVD
Sbjct: 254 EILKQFSRIPVYDVDDKKIIGLITLHYKLLSSEGRQGGPDQINMRKMLQELFVFPETIFVD 313

      Query: 301 DLLKALRNTQNMQMAILDEYGGVAGLVLTEDLLEIVGEIDDETQKAEQVHIEGNTYI 360
      DLL+ L RNTQNMQMAILDEYGGVAGLVLTEDLLEIVGEIDDETQ ASQV EI +N YI
40 Sbjct: 314 DLLRQLRNTQNMQMAILDEYGGVAGLVLTEDLLEIVGEIDDETQKAEQVHIEGNTYI 373

      Query: 361 VVGTMFLNEFNDFYFTELESDDVDVTIAGFYLTGVSIPQAEKAYEIVDSKKHITLIND 420
      V+GTMFLNEFNDFY+TELESDDVDVTIAGFYLTG+G+IP+QE+K AYE+D+KDH+ LIND
45 Sbjct: 374 VVGTMFLNEFNDFYFTELESDDVDVTIAGFYLTGIGTIPQEQKEAYEIVDSKKHITLIND 433

      Query: 421 KVKDGRITKLLKLSNIRQIIEE 444
      KVKDGRITLK++LS+IEQ IE+D
50 Sbjct: 434 KVKDGRITKLLKLSNIRQIIEE 457

```

SEQ ID 3718 (GBS70d) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 8-10; MW 65kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 11 & 12; MW 44kDa) and in Figure 179 (lane 5; MW 35kDa).

GBS70d-His was purified as shown in Figure 231, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1343-

Example 1195

A DNA sequence (GBSx1271) was identified in *S.agalactiae* <SEQ ID 3721> which encodes the amino acid sequence <SEQ ID 3722>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1212 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB84230 GB:AL162754 hypothetical protein NMA0960 [Neisseria
meningitidis Z2491]
Identities = 80/184 (43%), Positives = 119/184 (64%), Gaps = 3/184 (1%)

Query: 1 MIKRPILSHDFLAEVIDKEAITLDATMGNGNDTVFLAKSSK--KVYAFDIQERAIKNT 57
++K + +H L + + + LD T GNG+DT+FLA+++ KV+AFDIQ ++A T
Sbjct: 2 LLKNILFFAHCLLRQALPBGKGNALDGTAGNGHDTFLAQTAGIRGKWFADIQPQALNNT 61

Query: 58 KAKLTEQGISNAELILDHSHENLQYVHTPLRAAIFNLGYLPSADKIVITKPHITIKAIKN 117
++L E G SN LILDHSHENL+QY+ PL AAFIN G+LP DE++ T+ T+I A+
Sbjct: 62 RCHLEAGYSNVRLLLDHSHENLQYIPLDAAIFNPGWLPGGDKSLTIRTETSIALLSA 121

Query: 118 VLDILEVGGRLSIVVYGHGDKSEKDAVIAFVEQLPQNNFATMLYQPLQVNTPPFLIM 177
L +L+ G L ++Y GH+ GK E +A+ + + LPQ FA + Y N+ N+PP+L+
Sbjct: 122 ALSLLKENGMLIAVLVPGHENGKQBARAIEQWAKNLPQEQFAVLRYSTPIRNSPPVLLA 181

Query: 178 VEKL 181
EKL
Sbjct: 182 FEKL 185

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3723> which encodes the amino acid sequence <SEQ ID 3724>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.44 Transmembrane 127 - 143 ( 123 - 143)

----- Final Results -----
      bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9101> which encodes the amino acid sequence <SEQ ID 9102>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.44 Transmembrane 118 - 134 ( 114 - 134)

----- Final Results -----
      bacterial membrane --- Certainty= 0.157 (Affirmative) < succ>
      bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 124/184 (67%), Positives = 156/184 (84%)

```

Query: 1 MIKRPILSHDFLAEVIDKEAITLDATMGNGNDTVFLAKSSKVVYAFDIQERAIKNTKAK 60
N+KRPTILSHDFLAEV+DK ++ +DATMGNGNIT FLA+ +KKVYAFD+QE+AI KT +

```

-1344-

Sbjct: 10 MLKRPILHSHDFIAKVVVKSSVVVDATMGNGNUTAFIAQLAKKVVAFDVOEQAIRKTSRR 69

Query: 61 LTEQGISNAELLIDGHEMLSQVHTPLRAAIFNLGYLPSADKTVITKPHHTIKAIKINVL 120
L + G+NAELLIL GHE ++QVV P+RAAIFNLGYLPSADK++IT P+TT++A+ +L

5 Sbjct: 70 LAQLGLSNAELLILAGHEAVDQVITPEVRAAIFNLGYLPSADKSIITLFTNTLQALSKILT 129

Query: 121 LLEVGRGLSLMVYTHGIDGGSEKDAVIAFVBOLEPQNFAFMLYOPINQUNTPPFLIMVEK 180
+L VGRG++MYYTHGIDGG EKDA++ FV+QL Q + MLYOPINQUNTPPFLIM+EK

10 Sbjct: 130 LLMVGRGRIAMVYTHGIDGGSLKDALD FVKQLDQRKVSAMLYQFINQUNTPPFLIMLEK 189

Query: 181 LQSY 184
L +

Sbjct: 190 LADF 193

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1196

A DNA sequence (GBSx1272) was identified in *S. agalactiae* <SEQ ID 3725> which encodes the amino acid sequence <SEQ ID 3726>. Analysis of this protein sequence reveals the following:

20 Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1948 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AA00380 GB:AF008220 YtqA [Bacillus subtilis]
Identities = 161/302 (53%), Positives = 220/302 (72%), Gaps = 4/302 (1%)

Query: 2 KRYRAINDYYRELFGEKIFKLPIDAGFDCPNRDGTVARGGCTCTVSGSGDAVAPEAP 61
+KRY +N +RE FG K+EK+ +D GFDCPNRDGTVA GGCTFCA +GSGD

35 Sbjct: 13 EKRYHTILNYHLREHPGHKVFVKALDGGFDCPNRDGTVAHGCTCFCSAAGSGDFAGNITD 72

Query: 62 IREOPYKEIDFMHRKNPEVKNLYVFQNFPTNHAKLEIKERYEGAINPEGVIGINIGTR 121
+ QP+ +MH KW + KY+ YRQ PTNTHA +E++R+E + V+GI+I TR

40 Sbjct: 73 LITQPHDIKNRMHEKNKD-GKYIYFQAPTNTHAPEVLEKRFESVLAIDDDVGIATR 131

Query: 122 PDCLPDETIYYLAELSERMHVLELGLQTTYEATSA LINRAHSYDLYKKTVKRIRSLAPK 181
PDCLPD+ +YLAEL+ER ++ +ELGLQT +E T+ LINRAH ++ Y + V ++R+

45 Sbjct: 132 PDCLPDVVYDYLAELEHRTYLAELVGLQTVHERTALLINRAHDFNCTYVEGVNKLRLG-- 189

Query: 182 VEIVSHLINGLPGETHDMVKNVRRCVTNDIQGLKLHLHMTNTMQRDHYHGRLELL 241
+ + SH+INGLP E D++E + V D+QGK+HLLEH+ T M + Y +G+L L

50 Sbjct: 190 IRVCSHILINGLPLEDRDMMETAK-AVADLVQGIKTHLHLKSTPMVQKYEKGLKFL 248

Query: 242 SQDYISIIICDQLIIPKHIHVIHRTGDAFPHMLIGPMWSLNKWEVLNAIDKEMKRSY 301
SQ+DY++ ++CDQLSITP ++HRTGD P ++IGPMWS+NKWEVL AI+KE+E R SY

Sbjct: 249 SQDNYVQLVCDQLIIPKHIHVIHRTGDFELMIGPMWSLNKWEVLNAIDKEMKRSY 308

Query: 302 QG 303
QG

55 Sbjct: 309 QG 310

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3727> which encodes the amino acid sequence <SEQ ID 3728>. Analysis of this protein sequence reveals the following:

Possible site: 57
>>> Seems to have no N-terminal signal sequence

60

-1345-

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2023(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 260/307 (84%), Positives = 290/307 (93%), Gaps = 1/307 (0%)

Query: 1  MKKRYRAINDDYRELFGKIFKLPIDAGFDCPNRDGTVARGGCTFCTVSGSGDAIVAPFA 60
      MKKRY+ +N++YR+LFG K+K+K+PIDAGFDCPNRIGTVA GCTFCTVSGSGDAIVAPFA
Sbjct: 7  MKKRYQTLNEHYRQLPGAKMFKVFDIAGFDCPNRDGTVARGGCTFCTVSGSGDAIVAPFA 66

Query: 61  FIKBQFYKEIDFMHRKWPENVKYIVPQNFINTHAKLEIKRYEQAINSPGVGINIGT 120
      PI+ BQFYKEIDFMHRKWP+VN+YLVYPQNFINTH ++I++RYEQAINSPGV+GINIGT
Sbjct: 67  FIKBQFYKEIDFMHRKWPENVKYIVPQNFINTHDTVDVIRDRYEQAINSPGVGINIGT 126

Query: 121  RPDCLPDETIYYLAELSERMNVTLGLQITTYEATSLINRAHSYDLYKTKVKRIELAP 180
      RPDCLPD+TI YLAELSERMNV+ELGLQITTYE TS LINRAHSYDLYK+TV+R R P
Sbjct: 127  RPDCLPDITIAELSERMNVTLGLQITTYEETSLINRAHSYDLYKSTVRLRHY-P 185

Query: 181  KVEIVSHLINGLPGETHMMVENVRRCVTDNDIQIGIKLHLLHLMINTRMQRDYHESRLRL 240
      + IVSHLINGLP ETHMM+ENVRRCVTDNDIQIGIKLHLLHLMINTRMQRDYHESRL+L
Sbjct: 186  NNTIVSHLINGLPKETHMMLENVRRCVTDNDIQIGIKLHLLHLMINTRMQRDYHESRLKL 245

Query: 241  LSQEDYISIIICDQLEIIPKHIVIRITGDAFRHMLIGPMWSLNKWEVLNAIDKBEKQGS 300
      LSQ+DY+SIICDQLEIIPKHIVIRITGDAFR MLIQPMWSLNKWEVLNAIDKBE+R S
Sbjct: 246  LSQGDYVSIICDQLEIIPKHIVIRITGDAFRMLIGPMWSLNKWEVLNAIDKBEKRRGS 305

Query: 301  YQGCXAE 307
      +QGCX +
Sbjct: 306  FQGCXVD 312

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1197

A DNA sequence (GBSx1273) was identified in *S.agalactiae* <SEQ ID 3729> which encodes the amino acid sequence <SEQ ID 3730>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
40  INTEGRAL Likelihood = -9.82 Transmembrane 10 - 26 ( 6 - 30)
    INTEGRAL Likelihood = -4.73 Transmembrane 93 - 109 ( 87 - 112)
    INTEGRAL Likelihood = -4.57 Transmembrane 163 - 179 ( 161 - 181)
    INTEGRAL Likelihood = -2.97 Transmembrane 189 - 205 ( 185 - 205)
    INTEGRAL Likelihood = -1.97 Transmembrane 58 - 74 ( 58 - 74)
45  INTEGRAL Likelihood = -0.75 Transmembrane 130 - 146 ( 130 - 146)

```

```

----- Final Results -----
bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]
Identities = 62/159 (38%), Positives = 92/159 (56%), Gaps = 3/159 (1%)

55  Query: 34  ISFDCTIQSRVSGQLPNLSFRFFKLTIVIGNTVSQIAIAIMSVTFCT--LKKWYQARFI 91
      + FD+ + V+G L T K T IG+T S I +++ + F Y LK F
Sbjct: 34  LKFDVDVSLVQGRSPFLTDIMKFFTYIGSTASLILSLVLVFLPLIKRLRLVLELFT 93

60  Query: 92  AVNALISGICILSLKLIQKRVPTLHLVAGGYSFFSGHSMGTWFMGSIIILLQYTF 151
      AV + S + L +KL PQR RP L L+ GGYSFFSGH+M F +G + LL ++

```

-1346-

Sbjct: 94 AV-MVGSPLINLNVKLFQQRARPOLHRLIDIGGYSFPGSHAMAFSLYGLITFLWRHIT 152

Query: 152 KSIWKLQCGITGLLFLTGLSRIVLVGHVFPDVLVAGFI 190
+L L+I IG+SRIVLVGH+P+D+AG+G

5 Sbjct: 153 ARWARILLILFPMILSLIGISRIYLVGHVHYPSDIAGYL 191

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1851> which encodes the amino acid sequence <SEQ ID 1852>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -11.30 Transmembrane 154 - 170 (150 - 181)
INTEGRAL Likelihood = -10.88 Transmembrane 65 - 81 (58 - 93)
INTEGRAL Likelihood = -8.97 Transmembrane 10 - 26 (5 - 31)
INTEGRAL Likelihood = -3.77 Transmembrane 86 - 102 (86 - 105)
15 INTEGRAL Likelihood = -2.71 Transmembrane 185 - 201 (183 - 202)
INTEGRAL Likelihood = -1.54 Transmembrane 130 - 146 (130 - 148)
----- Final Results -----
bacterial membrane --- Certainty=0.5522 (Affirmative) < succ>
20 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/197 (44%), Positives = 134/197 (67%), Gaps = 1/197 (0%)
25 Query: 1 MLSRNSKLICAFIALILFSLGLVVIKYWPDVVISFDOTIQESVRGQLENLSTFFFKLAT 60
M +Q LI +F A++F +G +K++P+ + D TIQ +RG LP + T+FF+ +T
Sbjct: 2 MTNKQTHFLIASF-ALLIFVIIGYTVKFFPERLALLDNTIQAIRGNLPIVLTQFFRGVT 60
30 Query: 61 VIGNTVSQALAIMSVTFYLAGVYVQARFIANNAISGICILSKLIPQVRPRTIHLV 120
V GN ++Q+ + I+SV + KW +A FI N I+ I +LKL +QR RP + HLV
Sbjct: 61 VFGNVTQVLLVIVSVLVLFPMKWKIKIALFILNGAIAAFILTTKLKLVQRPFAIHLV 120
35 Query: 121 FAGGYSPSGHSMYTFMIFQSI ILLQYMPKSIWKLQCGITGLLFLIGLSRIYLVGH 180
+AGGYSPSGH+MG+ +IFGS++I+ + + + + +LI LIGLSRIYLVGH
Sbjct: 121 YAGGYSPSGHAMGMLIFGSLILCYORLSKLIQFVTSMIFIIILLIGLSRIYLVGH 180
Query: 181 FPDVLVAGFILAYGILN 197
+P+D+LAGF+L +GIL+
40 Sbjct: 181 YPSDILAGFVLGFGILH 197

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1198

45 A DNA sequence (GBSx1274) was identified in *S.agalactiae* <SEQ ID 3731> which encodes the amino acid sequence <SEQ ID 3732>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -8.44 Transmembrane 35 - 51 (33 - 59)
50 INTEGRAL Likelihood = -6.53 Transmembrane 193 - 209 (179 - 211)
INTEGRAL Likelihood = -4.46 Transmembrane 64 - 80 (60 - 82)
INTEGRAL Likelihood = -4.09 Transmembrane 108 - 124 (103 - 128)
INTEGRAL Likelihood = -2.71 Transmembrane 150 - 166 (148 - 166)
INTEGRAL Likelihood = -0.06 Transmembrane 174 - 190 (174 - 190)
55 ----- Final Results -----
bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
60 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1347-

A related GBS nucleic acid sequence <SEQ ID 9977> which encodes amino acid sequence <SEQ ID 9978> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAC83944 GB:L47648 putative [Bacillus subtilis]
    Identities = 53/186 (28%), Positives = 109/186 (58%)

Query: 33  RKMVTIAILSALSFLVMVVSFLIPGAFLKVDPSILPMLVAFILFDLKSSYGVLLRSL 92
           +K+V ++LS+++FVIM++FP ++LK+DFS +P +A+ ++ + V +++
Sbjct: 4   KKLVVVSMSSIAFVLMMLNFFPGFLPDYLKIDFSDVPAITAILTYGFLAGIAVRAIKNV 63

10 Query: 93  LKVILANRGPETFIGLEMMVAIALFLASFAIPWNRESAKDFIKASLFGTVSLTVMVA 152
           L+ I+ +G N +A LF+ A +K SAK + L GT ++T+ M
Sbjct: 64  LQYIIQSSNAGVPVGGQVANFIAGTLFILPTAFLEPKKINSAGKLAWSLLGTAAITILMS 123

15 Query: 153 LNTVFAIPLYAIFANFDIRKTFIVGNYLLTMVIPFNIVHGILISIVFYLT+VACLPILER 212
           LNTV +P Y F+ + + ++ ++PFN+++GI+I+VVF L ++ P +E+
Sbjct: 124 LNTVLLIPAYTWFLHSPALSDSALKTAVVAGILPFNMIKIGIVTVFSLIFIKLKPWIEQ 183

Query: 213 YKKTNV 218
           + ++
20 Sbjct: 184 QRSARI 189

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3733> which encodes the amino acid sequence <SEQ ID 3734>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 26
    >>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -6.48 Transmembrane 82 - 98 ( 74 - 100)
INTEGRAL Likelihood = -3.93 Transmembrane 161 - 177 ( 152 - 178)
30 INTEGRAL Likelihood = -3.61 Transmembrane 108 - 124 ( 107 - 126)
INTEGRAL Likelihood = -3.61 Transmembrane 33 - 49 ( 31 - 50)

----- Final Results -----
bacterial membrane --- Certainty=0.3590 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

40 >GP:AAC83944 GB:L47648 putative [Bacillus subtilis]
    Identities = 46/182 (25%), Positives = 97/182 (53%)

Query: 3   KTKHMIMIGILSAISFLMLVSPFAIIPGAFLKIEFSIIPVLEGLMINDKSAYLILMLR 62
           K K++++ +LS+I+F+LML++F +LKI+FS +P + +I + + ++
Sbjct: 2   KVKKLVVVSMSSIAFVLMMLNFFPGFLPDYLKIDFSDVPAITAILTYGFLAGIAVRAIK 61

45 Query: 63  SLKLFLNLRGVNDFIGLEMMIIATLAFVTAFLVWNRKQLSQVVFASLGTGLTGMN 122
           ++L+ + +G N IA LF+ A ++ + + LLCT +T N
Sbjct: 62  NVLQYIIQSSNAGVPVGGQVANFIAGTLFILPTAFLEPKKINSAGKLAWSLLGTAAITILM 121

50 Query: 123 VVLNNTFAIPLYAIFANFDIRAYIGVTKYMMVMVIPPFLVSEGLIPAITFYFYVIASKPIL 182
           +LNY +P Y F+ + + ++ ++PFN+++G++ + F ++I KP +
Sbjct: 122 SILNVTLLIPAYTWFLHSPALSDSALKTAVVAGILPFNMIKIGIVTVFSLIFIKLKPWIE 181

Query: 183 ER 184
           E+
55 Sbjct: 182 EQ 183

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 110/185 (59%), Positives = 144/185 (77%)

60 Query: 29  MNIIRKNTIAILSALSFLVMVVSFLIPGAFLKVDPSILPMLVAFILFDLKSSYGVLL 88
           N+ T KM+ I ILGA+SF+LM+VSF +IRGA FLK++FBI+P+L ++ DLKB+Y +LL

```

-1348-

Sbjct: 1 MSKTHKMINIGILSAISFLMLVSAIINGAFLKIEPSIIIPVIFGLMIMDLKSAYLIL 60

Query: 89 LRSLLKVLANRGPEFTFGLPMNVALALFLASPAIFWKNRESAKDFIKASLPGTVSLTV 148
LRSLLK+ L NRG FGLPMN++A+ALF+ +PA+ W +++ ++ ASL GT LT

5 Sbjct: 61 LRSLLKVLANRGVNDFTGLPMNIIAIALFPTAPALWNRQKTLGQVVFASLGLTGLTTF 120

Query: 149 SMVALNYVFAIPLVAIFANFDIRTFIGVGNVLIATMVPENIVEGILISIVFYLYVACL 208
MV LNY FAIPLVAIFAN DIR +IGV Y++TMVIPFN+VSG++ +I FY Y+A P

10 Sbjct: 121 GMVVLNYTFAIPLVAIFANIDIRAYIGVTKMMTMVIPFNIVEGLIFAITFYFVYIASKP 180

Query: 209 ILERY 213
ILERY

Sbjct: 181 ILERY 185

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1199

A DNA sequence (GBSx1275) was identified in *S.agalactiae* <SEQ ID 3735> which encodes the amino acid sequence <SEQ ID 3736>. Analysis of this protein sequence reveals the following:

- 20 Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -11.04 Transmembrane 278 - 294 (270 - 298)
- 25 ----- Final Results -----
bacterial membrane --- Certainty=0.5416 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 30 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3736 (GBS150) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 7; MW 29.7kDa) and in Figure 175 (lane 4 & 5; MW 30kDa).

Purified GBS150-His is shown in Figure 110A, Figure 199 (lane 5) and Figure 227 (lanes 6-7).

- 35 The purified GBS150-His fusion product was used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 110B), FACS (Figure 110C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1200

A DNA sequence (GBSx1276) was identified in *S.agalactiae* <SEQ ID 3737> which encodes the amino acid sequence <SEQ ID 3738>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

- 45 Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -15.34 Transmembrane 264 - 280 (257 - 285)
INTEGRAL Likelihood = -7.64 Transmembrane 23 - 39 (12 - 41)
- Final Results -----

-1349-

bacterial membrane --- Certainty=0.7135(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AA13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 95/271 (35%), Positives = 139/271 (51%), Gaps = 16/271 (5%)

10 Query: 29 VGLLITSYPPFNNWYNNKANNQVNFDMQIQKLMYKINRRFSLAKAYNRLDPSRLSD 88
 +GLL +YP ++W ++ Q ++ + E A AYN L + + +
 Sbjct: 1 NGILL--TYPTASWVSQYNSQVKTADYSAGVDGARP-DAKTQVQAHAYNDALSAGAVLE 57

15 Query: 89 FYTE-----KEKGIAEYAHMLEIAE--MIGYIDIPSIKQKLPYAGITTSVLEKAGH 140
 K +YA++L+ ++ + IPSI LP+Y GT L KG GH
 Sbjct: 58 ANHHVPTGAGSSKDSGLQYANILKANNBGLMARLKIPISLDLPVYHGTADDTLLKGLGH 117

Query: 141 LBTSLPIGKSSHTFVITAHRLGPKAKLPDLDLKKKGIPIYHNIKEVIAVKQDISVV 200
 LBTSLP+GG++ +VIT HRGL +A +PT+LDK+G + EVL Y+V VV
 20 Sbjct: 118 LBTSLPVGSEGTSTVITGHRGLAEATNPTNLKVKTGDSLIVEVFGVILTYRVTSTKV 177

Query: 201 KPDPISKLVLVVGKDYATLLCTPTYSINSHLLVRGRIKYVPPVKKEYNLAKLQTHYK 260
 +P+ L V +GKD TL+TCTP IN+HR+L+ G RI Y P K+ K
 Sbjct: 178 EPEETEARLVEBGLDILLVTCPLGINTRHILITGERI--YPTAKDLAAGKRDPVPHF 236

25 Query: 261 LYPILLILVILLVALL-----YLKRPKER 287
 ++ + + LI+V L L Y + KER
 Sbjct: 237 PWWAVGLAAGLIVVGLYINRSGYAAARAKER 267

- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3739> which encodes the amino acid sequence <SEQ ID 3740>. Analysis of this protein sequence reveals the following:

Possible site: 49

- >>> Seems to have no N-terminal signal sequence
 35 INTEGRAL Likelihood = -14.01 Transmembrane 225 - 241 (220 - 248)
- Final Results -----
 bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- >GP:AA13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 94/250 (37%), Positives = 133/250 (52%), Gaps = 17/250 (6%)
- 45 Query: 1 VECYKDRQLLSTHYKQVTPQKPSSEBEEVWQKAKAYNARLGIQPVDF-----SPFD 52
 V Y ++ + Y QV +P +V ++A AYN L V +A S +D
 Sbjct: 13 VSQYNSQSVKTADYSAGVDGARPDARTQV-EQAHAYNDALSAGAVLEANNHVPTGAGSSKD 71
- 50 Query: 53 GIHDQYVESLIQIENNIDMGVVEVPSIKVTLPIYHYTTDEVLTGAGHGLGALPVGSD 112
 Y ++L+ N +M +++PSI + LP+YH T D+ L KG GHG G+LPVGG+G
 Sbjct: 72 S--SLQYANILKANNBGLMARLKIPISLDLPVYHGTADDTLLKGLGILGDSLPVGSBC 129
- 55 Query: 113 THVTISAHRLPSAEMPTNLNLVKKGDTFYPRVINKVLAYKQDILTVBEDQVTSLSGVN 172
 T +VI+ HRGL A MPTNL+ VK GD+ V +VL Y+V VEP++ +L
 Sbjct: 130 TRSVITGHRGLAEATNPTNLKVKTGDSLIVEVFGVILTYRVTSTKVPEETEARLVES 189
- 60 Query: 173 GKDYATLIVTCITPVGVNTKRLVGRHRIAYHYKKYQAKAMKLVDSRWASVCAARGV 232
 GKD TLVITCTP G+NT R+L+ G RI Y K + K A G+
 Sbjct: 190 GKDLILVTCITPLGINTRHILITGERI-----YPTAKDLAAGKRDPVPHFPWWAVGL 243
- Query: 233 VIAILLVEMY 242
 +I+V +Y

-1350-

Sbjct: 244 AAGLIVVGLY 253

An alignment of the GAS and GBS proteins is shown below.

Identities = 93/192 (48%), Positives = 130/192 (67%), Gaps = 2/192 (1%)

5 Query: 52 VTNPDNCTQKLNTEKINRRFELAKAYNRTLDPSRLSDPYTEKKKKGIARYAHMIEIA--E 109
 ++ + Q + E+ ++ AKAYN L + D ++ + Y +L+I +
 Sbjct: 10 LSTYHKQVTQKKPSEMEEVWQKAKAYNARLGIQVPVDPAPSPRDGIHDKNYESLLQIENNID 69

10 Query: 110 MIGYIDIPSTIKQKLPYAGTSSVLEKGAHLEBTSLEPTGGKSSHTVITAHRLGPKAKLF 169
 +GY++PSIK LPYI TT VL KGAHGL G+LP+GG +HVI+AHRLG L A++F
 Sbjct: 70 IMGYEVVPSIKVITLPYHYTTDEVLTGKAGHLPGSALPVGDCGTHTVLSAHRGLPSAEMF 129

15 Query: 170 TDLKLLKGGKIPYIHNKEVLAYKVDQISVVKPDNPSKLLVVGKQYATLLTCTPPYSINS 229
 T+L+KKG FY + +VLAYKVDQI V+PD + L V GKDYATL+TCTPPY +N+
 Sbjct: 130 TNLNLVKKGDTFYFRVLMKVLAYKVDQITVEPDQVTSLSGVWKGQYATVCTPPYGVNT 189

Query: 230 HRLLVGRGHRIKY 241
 RLLVRGHRI Y
 20 Sbjct: 190 KRLLVGRGHRIAY 201

SEQ ID 3738 (GBS210) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 3; MW 61kDa).

GBS210d was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 2-4; MW 54kDa) and in Figure 187 (lane 9; MW 54kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 2-4; MW 28.7kDa) and in Figure 182 (lane 13; MW 29kDa). Purified GBS210d-GST is shown in lane 4 of Figure 237.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1201

A DNA sequence (GBSx1277) was identified in *S. agalactiae* <SEQ ID 3741> which encodes the amino acid sequence <SEQ ID 3742>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

35 Possible site: 42
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.61 Transmembrane 20 - 36 (15 - 40)
 INTEGRAL Likelihood = -7.27 Transmembrane 259 - 275 (258 - 277)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.5246 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 76/219 (34%), Positives = 120/219 (54%), Gaps = 12/219 (5%)

50 Query: 28 LSTLLYFVVSFRFYTTIESNNQTOQFERAAKXLSCKEINRRMALAQAYNDSLNI-----N 80
 +L YP + + + T D+ A ++ + + A AYND+L+ N
 Sbjct: 1 MGLLTYPITASWVSQYNQSKVTADYS-AQVDGARPDAKTQVBOAHAYNDALSGAGVLEAN 59

Query: 81 VHEDEPYEKKRIQKGVABYARMLEVSEK--IGTISVPIQKQKLPFAGSSQEVLSKQAGH 138

-1352-

EVFPGVLYZYVSTVSTKVEPRSTEARVVEGKGLATVYCTPLGINTTRILLNGERKI-YPTPAKD-LAAAGKRPDVPVFPW
 170 180 190 200 210 220 230
 Y-----
 1098 1179 1209 1239 1269 1299
 I-----ISLVIIIANLWL--IKRQKQNR-LASVRKIGS+WEENPKTKLNSRF+IDG+MA+YYS+LIV+*PHILF+
 WAVGLAAGLIVLYLGRSGSAAARAKERALARARAAQSPQPTWABQRIRIMDDGAGPDRPTLVLPVQPPQSPREN
 250 260 270 280 290 300 310

SEQ ID 8750 (GBS212) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 4; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 2; MW 61kDa).

Purified Thio-GBS212-His is shown in Figure 244, lane 5.

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1202

A DNA sequence (GBSx1278) was identified in *S. galactiae* <SEQ ID 3743> which encodes the amino acid sequence <SEQ ID 3744>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 29
    >>> Seems to have a cleavable N-term signal seq.
        INTEGRAL Likelihood =-10.40 Transmembrane 680 - 696 ( 674 - 699)

    ----- Final Results -----
25 bacterial membrane --- Certainty=0.5161(Affirmative) < success
    bacterial outside --- Certainty=0.0000(Not Clear) < success
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the GENPEPT database.

```

30 >GP:CA5745>|GS:X1869 crf2 |Lactobacillus leichmannii|
  Identities = 84/325 (25%), Positives = 122/325 (36%), Gaps = 94/325 (28%)

Query: 397 VNVV+LTKDKD-----KTVASVSLTKSGTGT---DLGNGIKFEVSGNF 437
      VNV + +DKD          TV+ LTK+ + T+ D G + P+ +
35 Sbjct: 236 VNVFNNKDKDPTFNVVDKPTDGDIDASTVSDGLKTSDDTVKNGIKYQVFKT--- 292

Query: 438 SGKPGLENGEYVTSERVSYGSAINLENGKATITNTKDSNDFPLTPEPKVETGKFK 497
      S L K S I+          K T+T+          + + T+ +G+
40 Sbjct: 293 SAVALQKGLSITTY-----KATLTWNATPDKA---IGNTATSLGWTGNI 336

Query: 498 VKINQQDRL--AGAGFVVISAGHYLAKADSGEGQKTLAAKKIALDRAALVYNKSLAT 555
      G R+  GAQFV K+S          QSKL+ L+ + N+N+
50 Sbjct: 337 TSTPANGPRITGGAGFVKKDS-----GKMTLAGAEPLQVDSNGNIVSYA 384

Query: 556 DKSGEIKTAKELIKTKQADYDAAFIEARTAYEWITDKARATITYSNDGQFEVTLADG 615
      Q +          +Y W + A TTYS+ G + GL+
45 Sbjct: 385 TQASDG-----SYWNDSATFATITYSNDGLVALKGLGYS 420

Query: 616 -----TYNIERTLAPAGFAKLAGNIKFVNQGSYITGSDIVDANSQKQDTRVENKK 668
      +Y L E AD GAKL +KF + QGS- G+ + + N K+
50 Sbjct: 421 DKLDGSSVALLGTAQPCYAKLDSEPVKFSITQSF--GDSNKITIDNTKSG-- 470

Query: 669 VTIPGTGGIIFLFTIIGLSIMLGA 693
      +P TGG G +F IG+ I+ A
55 Sbjct: 471 -LIPSTGGKGIYFLAIGIVIMVA 494

```

No corresponding DNA sequence was identified in *S. pyogenes*.

-1353-

SEQ ID 3744 (GBS59) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 8; MW 120kDa), in Figure 11 (lane 9; MW 100kDa) and in Figure 13 (lane 6; MW 74kDa).

GBS59-His was purified as shown in Figure 193, lane 2.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1203

A DNA sequence (GBSx1279) was identified in *S. agalactiae* <SEQ ID 3745> which encodes the amino acid sequence <SEQ ID 3746>. Analysis of this protein sequence reveals the following:

10 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.13 Transmembrane 870 - 886 (864 - 887)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.2253 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAD33086 GB:AF071083 fibronectin-binding protein I [Streptococcus pyogenes]
 Identities = 58/176 (32%), Positives = 83/176 (46%), Gaps = 19/176 (10%)
 Query: 6 KFSKILTLSLFCLSQIPLNINVLGEST---VPENGA--KGLVVKKTDDQNKPLSKATFV 60
 K S +L+L+ F L + + + G S NGA +G +KK D NKPL AT
 25 Sbjct: 8 KLSFLSLTGFILGLLLVFIGLSQVSVGHAEITRNANKQGSFEIKKVDQNKPLPGATSS 67
 Query: 61 LKTTAHPESKIEKVTAELTGEATFDNLIPGDYTLSEETAPEGYKKINQVQVKNESNGKT 120
 L + + ++ T+ G NL PG YTL BETAP+GY KT++TW V V NG T
 30 Sbjct: 68 LTSKDGKGI SVQTPTSNDKGI VDAQNLQPGTYTLKRETA PDGYDKTSRTWTVVYENGTY 127
 Query: 121 TIQNSGDINKSTIGQNSBELDKQYPPPTGIYEDTKESYKLEHVKSVPEN--GKSEAKA 174
 + + I + +D S +L+ K SV + GK+ E +
 Sbjct: 128 KLVENPYNGRIISKAGS-----KDVSSSLQLENPKMSVSVSKYKTEVSS 171
 Identities = 31/92 (33%), Positives = 49/92 (52%), Gaps = 14/92 (15%)
 35 Query: 725 PTTITIKNEKKLGRIFIKVDKNNKLLKLGATFELQRFNEDYKLYLPKNNKSVTVGRN 784
 P+IT+ N K++ ++ P K+ DN + L A FEL+ N+ K+ N
 Sbjct: 501 PSITVANLKRVAQLRFKIMSTDN--VPLPEAAFEIIRSSN-----GNSQKLEASN 548
 40 Query: 785 --GKISYKDLKDGKYLIEAVSPEDYQKITNK 814
 G++ +KDL G Y L E +P+ YQ++T K
 Sbjct: 549 TCGEVHFKDLTSGTYDLVETKAPKGYQGVTRK 580

No corresponding DNA sequence was identified in *S. pyogenes*.

- 45 SEQ ID 3746 (GBS67) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 10; MW 140kDa), in Figure 11 (lane 10; MW 150kDa) and in Figure 12 (lane 6; MW 95.3kDa).

GBS67-His was purified as shown in Figure 192, lane 10.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.
- 50

Example 1204

A DNA sequence (GBSx1280) was identified in *S.agalactiae* <SEQ ID 3747> which encodes the amino acid sequence <SEQ ID 3748>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.2020(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9979> which encodes amino acid sequence <SEQ ID 9980> was also identified.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3749> which encodes the amino acid sequence <SEQ ID 3750>. Analysis of this protein sequence reveals the following:

```

    Possible site: 58
    >>> Seems to have no N-terminal signal sequence
20  INTEGRAL Likelihood = -1.75 Transmembrane 393 - 409 ( 392 - 409)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 122/325 (37%), Positives = 186/325 (56%), Gaps = 5/325 (1%)

30  Query: 7  LIRNYLEKDIINQIKIATLCY--DYPSITIDKSKHGLISELLIRKYCHDITLTPNSQL 64
    LIE YLE I ++ +L+ L + Y P + + + GL+ L + YC +L F L
    Sbjct: 1  LIEKYLESSIESKQQLIVLFFKTSYLP---ITEVAEKTGLTFLQLAHYCEELNAFFPGSL 57

    Query: 65  SLNIEKSTIVYQSGVTVREQAFKIYHQSHVLQLLKFILTNDSGRLPITYFSEKFGSLCA 124
    S+ I+K I Q ++E ++Y S+VLQLL FL N S PLT F+ L$ +
35  Sbjct: 58  SNTQKRMISQCFTHPFKETYLYQLYASSNVLQLLAFLLKNGSHSRPLTDFASHPFLSNS 117

    Query: 125  TAYRIRKHISPLLEKLGQFQIVKNTITGDEYRIRYLIAPINAQFGIEVYPM$MDKLLIKR 184
    +AYR+R+ + PLL ++ KN I G+EYRIRYLIA L ++FGI+VY +++ DK I
40  Sbjct: 118  SAYRREALIPLLENFELKLSKNKIVGGEYRIRYLIALLYKFGIKVYDLTQDQKNTIHS 177

    Query: 185  LLLHSTTFAGHYFPNTFIFFDTLLSLSWKRINNVVVPYSSLFTELQNIPIYDTLQYC 244
    L ST S + +F F+D LL+LSWKR ++V +P + +F +L+ +F+YD+L+
    Sbjct: 178  FL$HSS$THLKT$PWL$H$S$P$FYDILLAL$WKR$H$Q$SVTIPQTRIPQ$K$K$LFVY$SLK$K$ 237

45  Query: 245  VKNVIIDSFKINKLKKDDIYIFLAVLTSHNSFENPNWTEKRIDNVIAIFENYPKFQKLLQ 304
    ++I ++N D+DY++L Y++NSF++ WT + I +FE F+ LL
    Sbjct: 238  SHDIETCYQLNFSAGDILVKNLYITANN$FASLQWTP$HIRQYQCLF$RENDTFRLLLN 297

50  Query: 305  PLKDALPLSGSYHDELIVKVAIFFS 329
    P+ LP LVK +PFS+
    Sbjct: 298  PIITLLENLKEQKASLVKALMPFSK 322

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1355-

Example 1205

A DNA sequence (GBSx1281) was identified in *S.agalactiae* <SEQ ID 3751> which encodes the amino acid sequence <SEQ ID 3752>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 21
   >>> Seems to have no N-terminal signal sequence (or aa 1-22)

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.1168 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AA899071 GB:U67549 galactosyltransferase isolog [Methanococcus
   jannaschii]
   Identities = 108/395 (27%), Positives = 196/395 (49%), Gaps = 28/395 (7%)

Query: 4   KVKTAVFSGGYLFLPGGIERYTDMTADLVK-RGYRVVIVTTNHGDLPIIDEDKGR--- 59
      K+K + +F GYY+P +GG+E +D+ T L + Y + I N +P E + R
20  Sbjct: 3   KIKLI-IPFGYIPIHGLETHVDEFTGHLSEDENYDIYFAPN---IPKYKEFIRHNN 58

Query: 60   -KIYKLETKNIVKQRYPIINK-NREYNTLMKYVSDENIDFVICNTRFQLTILEGLSPAKN 117
      K+YR P I+ YP+ N N++ + + + D V+ TRF TL G FAK
25  Sbjct: 59   VKYTRYPAFELIPN-YVPVNIPIFKWRFNPNLYKIDFDIVMTRTFPSNTLLGFTFAKL 117

Query: 118  HHLPS--IVLDHGSSRFSVNNRFLDPFGAIYEHLL/TARVKHYRDPFIYAVSKRSVEWLKHF 175
      I ++HGS+ + + F + Y+ + + A+SK ++
30  Sbjct: 118  RPKKKKLHVEHGSAPVKLESEFINKLSYFYDKTIGLIPKKADYVVAISKAVINFIEN 177

Query: 176  NIEAKGV--IYNSVS---ESLGSDFAGTAYLEKSADDIFITYAGRIKEKGIIELEAF 229
      + K + IY + ES+G D EK + I + + GR+ K KG+E +++A+
35  Sbjct: 178  FVND/CIPIIYRGLIEKEIKESIGED---KKIKFNNKIKLCFVGRLYRWGVENIKAY 234

Query: 230  S--MSQYSENVYLOIAGDGPFLAHLKE---KYQSKQINFLGKLNFBQYMSLNAQTDFVY 284
      E + L + G G +L LK+ Y + I P GK++FE+ +++ +DI+++
40  Sbjct: 235  VDLKDLKKEKILIVVGYGSDLERLKLKLAGNYLWNGIYFTGRVDFEKALAIYKASDIYIH 294

Query: 285  PSMYPBELFTSILEAGLSSAIIATDRGGTVEVIDSPELGIMBENT-QSLHESLDLLVK 343
      S GL +S+L+A AI+A+ G EV+ GI++++N+ + + + L++
45  Sbjct: 295  SSYKGGLSSSLLLQANCGKATVAPYEGADEVINDGYNGILLKDNSPEIKRGIKILIE 354

Query: 344  DKALREKLQNIARKIEHFTWEKTVKELDYIIQK 378
      + LR+ +N IKE+P W+K+V++ I ++
50  Sbjct: 355  MNLRKIYGENAKNFIKENPNWKKSVKEYKIFER 389

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3752 (GBSx258) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 2; MW 43kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 7; MW 67.9kDa).

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1206

A DNA sequence (GBSx1282) was identified in *S.agalactiae* <SEQ ID 3753> which encodes the amino acid sequence <SEQ ID 3754>. Analysis of this protein sequence reveals the following:

55 Possible site: 31

-1356-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1182 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP: CAB52237 GB: Z98171 EpsQ protein [Streptococcus thermophilus]
 Identities = 112/278 (40%), Positives = 163/278 (58%), Gaps = 2/278 (0%)

Query: 1 MKYLAGIVTFNENIERLDQNIIRAIYQVSHIYIVDNGSKNKEISQLVADYNEGHLITVD 60
 M AGIV FNF+I+RL +NI A+ Q +H+Y+VDNGS N +E+ L+ YN+ +++
 15 Sbjct: 1 MDISAGIVLNFDFIKLKENIDAVIIQCTHLYLVDSGSGNVDEVKLLINQNGS-KISIL 59

Query: 61 YLTENKGIAYALNCIGQFAVAQEFDFWFLTLQDQSVVLGDLIDNTYNIHLHPKVMGLSCLY 120
 + EN+GIA ALN + A + FDM LTLQDQSVV +++ +E Y++ V3+L +
 15 Sbjct: 60 WNRNQGIGIAKALNQLTSAQKESGDFWFLTLQDQSVVPSIVGFEFKYINNSVSGILCFII 119

Query: 121 QDMRENLMQEFDFYKIEECITSAALNKTSVVFETSGFABEMFIDFVDESNYRLSEMG 180
 D N++ + D EI+ECITS +L+ + E GF E MFID VD ++ YRL + G
 20 Sbjct: 120 CDNRNDEBIKINEDCTEIDSCITSGSLANIKANSBIOGFDERMFIDGVPDICTRLRQG 179

Query: 181 YKTYQVNFYGLLHIEHSGSRVKKFGHVHFIANSFPKTYMIRNATYLIKKYKKRKYK 240
 YK Y ++ + LLMR+GH + KK V NIS FRKY+ RN IY KK
 25 Sbjct: 180 YKTCIHSVLLLEHGHIEYHFLFKVLVQNHGAFFKYIARNIYTKAKRSTLNVVK 239

Query: 241 LVPMRNEFVRVLV-AEEQSKKIVAMIKLDGLMKV 277
 + + + + + EE K KI + +G+ DG KV
 30 Sbjct: 240 GLLQEKILGIVIFYEEDKLNKIRCICRGIYDGFPGKV 277

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 **Example 1207**

A DNA sequence (GBSx1283) was identified in *S.agalactiae* <SEQ ID 3755> which encodes the amino acid sequence <SEQ ID 3756>. This protein is predicted to be EpsU protein (rfbX). Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have an uncleavable N-term signal seq

40 INTEGRAL Likelihood = -8.44 Transmembrane 357 - 373 (352 - 387)
 INTEGRAL Likelihood = -7.59 Transmembrane 88 - 104 (79 - 107)
 INTEGRAL Likelihood = -7.32 Transmembrane 440 - 456 (433 - 465)
 45 INTEGRAL Likelihood = -6.48 Transmembrane 246 - 262 (245 - 263)
 INTEGRAL Likelihood = -4.78 Transmembrane 294 - 310 (290 - 312)
 INTEGRAL Likelihood = -3.88 Transmembrane 164 - 180 (162 - 183)
 INTEGRAL Likelihood = -3.56 Transmembrane 144 - 160 (136 - 161)
 INTEGRAL Likelihood = -2.87 Transmembrane 317 - 333 (316 - 334)
 50 INTEGRAL Likelihood = -2.71 Transmembrane 374 - 390 (374 - 393)
 INTEGRAL Likelihood = -0.96 Transmembrane 44 - 60 (44 - 62)
 INTEGRAL Likelihood = -0.80 Transmembrane 15 - 31 (15 - 32)

----- Final Results -----

55 bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB52225 GB: Z98171 EpsU protein [Streptococcus thermophilus]

-1357-

Identities = 189/462 (40%), Positives = 313/462 (66%)

Query: 1 MKLLRMFPNTSYQLLTLLPLVTVFVSRVLSPOQGINAVTSSIVMYFLFGALGISL 60
 M+++KH YN YQ+ +++PL+T+PY+SR+L P GIGIN+YT+SV YF LFC+G+ L
 Sbjct: 1 MQIVKYLNYALYQFIIIVPLLTIPYLSRILGPSGIGINSYTNSTYVQVFLVFGSIGLGL 60

Query: 61 YGNREIAPVQSNKYKRSKIPWELVVLKIASVSIATLLPFGFVLLTNEWQLPYLQGINLL 120
 YGNR+IAPV+ N+ K SK+P+E+ +L+L ++ A+ LF F+++ ++ +YL Q I ++
 Sbjct: 61 YGNRQIAPVEDNQVMSKVPTETIFLRFTI CLAYFLVAFPLIINGQYAYVLSQSIAIV 120

Query: 121 ATATDISWFFIGVEDFKIIVIRNTIVKLTIVVLTVLVKTPDOLALYMLFIAPASLLGNL 180
 A A DISW F+G+B+FK+IV+RM IVK+ + FL VK+ +DL +Y+ + ++L+GNL
 Sbjct: 121 AAADFISWAFMGIEKFRVIVLRNFIKLLALFSIFLVKSYNDLNIYILITVLSLIGNL 180

Query: 181 TVHHHLKHELIKIPPSRLDILIHAFPIIMFLPQITMQIYLSINKSMILGAMISVSAGYF 240
 T + L ++K+ + L + HL+ +L++P+PQI +QIY LNK+MLG++DSV S+G+F
 Sbjct: 181 TFFPSLHRYLVKVNKELRPIKHLKQLVNFIPQIALQITVWLKNTLGLSLDSVTSGGFF 240

Query: 241 DQSDKIIRIIFLTVSAIGQVFLPRSSLPSSGKQKAKALLKLVDSNAISMLIAGV 300
 DQSDKI++++ IV+A G V LPR+++ F+ + + K + ++AIS+ N+ G++
 Sbjct: 241 DQSDKIKVLVLAIVTATGTVMLEPRVANAFHREYSKIKEMTAGFSFVSATISIPMWGLI 300

Query: 301 GVSSTFAVFFFGKGYEAWGPLMAVEGLMIICISYGNALCTQVLLASRRYKATMSAVIGL 360
 ++ F FF + V P++ +ES+ II I++ NA+G QYLL + + K+YT+S +IG
 Sbjct: 301 AITPKFVLPFFTSQSDVIPVIMTESIATIFAMSNAIGNQYLLATPWKRSYTVSVIIGA 360

Query: 361 VANVVLNILLIPILGRGAILSTVITFVLSVQAISLSDVFTFKELRGMLRYLLAATL 420
 + N++LNI LI LGR+GA I+TVI+E V+YQ + L + +YLLA +
 Sbjct: 361 IVNLMNLPLIYLGAVGASIAIVISEMSVTVYQLFIHKQMLNHLTSPDLSKYLIAGLV 420

Query: 421 SGAVLYIYNTQMSVLSVNVVQSISAVITVYGVIVITKAPVI 462
 +++ I+ S + ++ V + IY+ ++ KA +I
 Sbjct: 421 MFLIVFKISLLTPTSWIFILLEITVGIITVILLIFLKAETI 462

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1208

40 A DNA sequence (GBSx1284) was identified in *S.agalactiae* <SEQ ID 3757> which encodes the amino acid sequence <SEQ ID 3758>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

45 ----- Final results -----
 bacterial cytoplasm --- Certainty=0.1742(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1209

55 A DNA sequence (GBSx1285) was identified in *S.agalactiae* <SEQ ID 3759> which encodes the amino acid sequence <SEQ ID 3760>. Analysis of this protein sequence reveals the following:

-1358-

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1210

A DNA sequence (GBSx1286) was identified in *S.agalactiae* <SEQ ID 3761> which encodes the amino acid sequence <SEQ ID 3762>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.56	Transmembrane	214 - 230 (210 - 236)
INTEGRAL	Likelihood = -10.03	Transmembrane	364 - 380 (361 - 386)
INTEGRAL	Likelihood = -7.96	Transmembrane	272 - 288 (271 - 291)
INTEGRAL	Likelihood = -6.95	Transmembrane	23 - 39 (20 - 41)
INTEGRAL	Likelihood = -5.57	Transmembrane	191 - 207 (189 - 209)
INTEGRAL	Likelihood = -5.15	Transmembrane	434 - 450 (425 - 451)
INTEGRAL	Likelihood = -4.25	Transmembrane	143 - 159 (138 - 162)
INTEGRAL	Likelihood = -3.13	Transmembrane	167 - 183 (166 - 186)
INTEGRAL	Likelihood = -1.44	Transmembrane	400 - 416 (400 - 416)
INTEGRAL	Likelihood = -1.33	Transmembrane	333 - 349 (333 - 349)
INTEGRAL	Likelihood = -0.80	Transmembrane	232 - 248 (232 - 251)

30 ----- Final Results -----

 bacterial membrane --- Certainty=0.5225 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1211

40 A DNA sequence (GBSx1287) was identified in *S.agalactiae* <SEQ ID 3763> which encodes the amino acid sequence <SEQ ID 3764>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.1792 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

A related GBS nucleic acid sequence <SEQ ID 9981> which encodes amino acid sequence <SEQ ID 9982> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAF18951 GB:AF155805 Cps9H [Streptococcus suis]
    Identities = 53/116 (45%), Positives = 75/116 (63%), Gaps = 4/116 (3%)

    Query: 6  VLMATYNGQGFIHQDLSIRNQTILRPDYVLMRDGSGTDDTVKVVEDYIKEHRLDGSWITS 65
              VLMATYNG FI QLDLSIRNQT++ D V++ DD STDVT+K+++DYIK++ LD W ++
    Sbjct: 4  VLMATYNGSPFIHQDLSIRNQSVSADKVIWDDCSTDTVTIKIKDYIKKYSLDGWVVSQ 63

10  Query: 66  NDKNLGRINFRQLLDVLAIEVDYVFFSDQDDTWYHKNKMQVDIMBERQDINLL 121
              N N G F L + VFFSDQDD W HK + I +R+++++
    Sbjct: 64  NKSNGHYQTFINL---TKLVQBGIVFFSDQDDIWDCCHKIETMLPIF-DRENVSMV 115
  
```

- 15 Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 1212

A DNA sequence (GBSx1288) was identified in *S. agalactiae* <SEQ ID 3765> which encodes the amino acid sequence <SEQ ID 3766>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

```

    Possible site: 13
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
25  bacterial cytoplasm --- Certainty=0.1278(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

- 30 A related GBS nucleic acid sequence <SEQ ID 9983> which encodes amino acid sequence <SEQ ID 9984> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF18951 GB:AF155805 Cps9H [Streptococcus suis]
    Identities = 57/146 (39%), Positives = 81/146 (55%), Gaps = 8/146 (5%)

35  Query: 10  VLMATYNGEIFISQSLIRNQTILKPDYVLLRDCSTDBETVNVNYSIAKHELBGKIVK 69
              VLMATYNG FI +QLDSIR Q++ D V++ DDCSTD+T+ ++ +YI K+ L+ W + +
    Sbjct: 4  VLMATYNGSPFIHQDLSIRNQSVSADKVIWDDCSTDTVTIKIKDYIKKYSLDGWVVSQ 63

40  Query: 70  NDKNLGRINFRQLLDVLAIEVDYVFFSDQDDIWLDNERQCFAIMSDKPKQIEVLSADV 129
              N N G F L + VFFSDQDDW K E I D + + + + V
    Sbjct: 64  NKSNGHYQTFINL---TKLVQBGIVFFSDQDDIWDCCHKIETMLPIF-DRENVSMV---V 115

    Query: 130  DFKWMSTEAQVPHFLTFSSSDRISQY 155
              K+ + + + +SDRI+ Y
45  Sbjct: 116  FCKGRLLDENGNISSPPTSDRINITY 141
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1213

- 50 A DNA sequence (GBSx1289) was identified in *S. agalactiae* <SEQ ID 3767> which encodes the amino acid sequence <SEQ ID 3768>. This protein is predicted to be dTDP-glucose 4-6-dehydratase (galE). Analysis of this protein sequence reveals the following:

-1360-

Possible site: 44

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.02 Transmembrane 250 - 266 (250 - 266)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1808 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 9985> which encodes amino acid sequence <SEQ ID 9986> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC14890 GB:AJ295156 d-TDP-glucose dehydratase [Phragmites

australis]

15 Identities = 108/327 (33%), Positives = 170/327 (51%), Gaps = 22/327 (6%)

Query: 29 ANKGVLSSNSMLASYMVELLAYLMNSTRVYQTLIATARNIEKARDKPSDLVCKDYPTL 88

AN +L+G + S+V L N + +I +D +G F L

20 Sbjct: 33 ANLRILVTGGAGFGTSHLVKIM-----ENKEHEVIVADNFFIGSKDLKKWIGHPRPEL 87

Query: 89 IPYDVEERLEYDGKVOYIIHAASNASPTAILSNPVSIIKANTIGTILNLDPAKERTIENF 148

I +DV + L + VD I H A ASP NPV IK N IGTLM+L AK +

Sbjct: 88 IRHDVTQPLIVE--VDQIYHLACASPFIYKPNPVKTIKTNWIGTLMGLAK-RVGARI 144

25 Query: 149 LFLSTREYVGTSIKEVIDEAYGGFOILATACYPESKRMETLLQSYTDQYKVPPTIAR 208

L ST EYVG ++ E +G + + R+CY E KR-AETL+ Y+ Q+ + IAR

Sbjct: 145 LLTSTSEYVGDPLEHPQTEAYWGNVNPICVRSYDEGKRVATLMDFYHRQHGIEIRIA 204

30 Query: 209 IAHSEFGPMELGNDGRIMNDLLSNVIDGKDIVLKSSGTAERAPCYLADAVSGLFTILLNG 268

I +++GP M + +DGR++++ + + G + ++ GT R+FCY+AD V GL L+NG

Sbjct: 205 IINTYGFPMNI-DDGRVSNFIAQAVRGDELTVQKPGTQTSTFCYVADMDGLIK-LANG 262

Query: 269 RVQAYNVANEDQPMIKDLQKINDLPSDKNISVFDIPKIMSAGYSKMRTR--LTM 325

N+ N + M+ +LA+K+ L + + + TM+ R R +T

35 Sbjct: 263 NITGFINLNGPGETML-ELAEVKELINP-----EVTVMIENTPDDPRQRKEDITK 314

Query: 326 AKLEALGKREVSLESGLTKVQAFEE 352

AK E LGW+ +V L G++ F E

40 Sbjct: 315 AK-EVLGMEPKVLRDGLVLMEDDFRE 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1214

45 A DNA sequence (GBSx1290) was identified in *S. agalactiae* <SEQ ID 3769> which encodes the amino acid sequence <SEQ ID 3770>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9987> which encodes amino acid sequence <SEQ ID 9988> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11866 GB:X99104 similar to hypothetical proteins [Bacillus subtilis]

-1361-

Identities = 77/231 (33%), Positives = 131/231 (56%), Gaps = 6/231 (2%)

Query: 13 VIFAGGVGRMRMTKGPFLVHGKPIIVHTIDIPONTEAIDAVVVVCSMDLYMNNL 72
 VI A G G+RM G+ K F+E+ G P+I+HT+ +F + D +++V ++ L
 5 Sbjct: 6 VIPAAGQGKRMKA-GRNELFTELKGDPEVIHTLRVDFSHRQCCKIILVINEQRREHFQQL 64

Query: 73 VERFNLTKVKAVVAGGFTGQMSIFKLEAAEQIATDDAVVLHDGVRPLINEEVINANTQ 132
 + + +VAGG+ Q S++KGL+A +Q + +VL+HDG RP I E I+ I
 10 Sbjct: 65 LSDYFPQTSIELVAGGDERQHSVYKGLKAVKQ---EKIVLHDGGRPFIKHEQIDELIA 120

Query: 133 SVKETGSAVTSVRAKETVVLNDSSKISEVVDRTSGFIKAPQSFYDLISLVERDAISK 192
 ++TG+A+ +V K+T+ V D ++SE ++R+ + + PQ+P LS ++ +A K
 Sbjct: 121 EAEQTGAAILAVPVKDTIKRVQDL-QVSETIERSSLAVVQTPQAFRLSLIMKAHAASERK 179

Query: 193 GITDAIDGSTLMGMYNRELITVGGPYENIKITTPDPYMFKALYDARENEQ 243
 G D+S + M + +VGG Y NIK+TTPDD +A+ ++ +
 15 Sbjct: 180 GFLGTDRAISLVEQMGSSVRVVGSGYINIKITTPDDLTSAEALMESESGNKK 230

No corresponding DNA sequence was identified in *S.pyogenes*.

20 SEQ ID 3770 (GBS647) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 9 & 10; MW 55.9kDa + lane 8; MW 27kDa) and in Figure 186 (lane 5; MW 56kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 12; MW 31kDa), in Figure 140 (lane 9; MW 31kDa) and in Figure 178 (lane 6; MW 31kDa).

25 Purified GBS647-GST is shown in Figure 243, lane 4; purified GBS647-His is shown in Fig.229, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1215

30 A DNA sequence (GBSx1291) was identified in *S.agalactiae* <SEQ ID 3771> which encodes the amino acid sequence <SEQ ID 3772>. This protein is predicted to be LicD1. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2647(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9989> which encodes amino acid sequence <SEQ ID 9990> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD37094 GB:AF106539 LicD2 [Streptococcus pneumoniae]
 Identities = 85/271 (31%), Positives = 130/271 (47%), Gaps = 15/271 (5%)
 45 Query: 1 MKEMTVSEIREVQLEMLAYIDKVARNDNKIEYSLGGGSLGAMRHKGFIIPWDDDDIDMLER 60
 M+ + EI+B+QL +L YID+ + + I Y L G++LGA+RHKG IPWDDDDID L R
 Sbjct: 1 MOYLEKKEIKETQLALLOYIDETCKKHIDIPYFLSYGTMIGAIRHKGMIIPWDDDDIDISLYR 60

50 Query: 61 SOYERILMKALDANNDSFKLLHBSVEKNL---FPAKLYHTKSMYLSKTORHPWIGFI 117
 YERL+K + + N+ +K+L S + + W FA + T + + T +FI
 Sbjct: 61 EDYERILKLIIEEENHPRYKVL--SYDTSWYFNFASILDTSVIEDHVKYKRHDTSLEFI 118

-1362-

Query: 118 DIFPLDRLPESAFERQRFKKVHSAANLACTTYPNFASGRKLYANAKLIIGLP-RPIA 176
 D+FP+DR ++ + + + A L G KL RL RP+
 Sbjct: 119 DVFPIDRPDLSIVDKSY---KYVALRQLAYIKKRAVHGDSKLDFLALCSWYALRPVN 175

5 Query: 177 YHGQAKKRRARTVDQVMETYNQEVPMGYTD-SRYRLKRYFPREI PSEYEDVMFENIKTR 235
 KK +EQ+++ Y G + +KE FP + F E FE
 Sbjct: 176 PRYPYKK---IDLQVKNVINTPQYBGGVGIGKEGKKEIFPVDTFKELILTEFEGRMPL 231

10 Query: 236 KINNEHAYLNQLYOGSYMELPPESKREHSY 266
 K +L Q+Y G YM P + +E +R+
 Sbjct: 232 VPKKYDQPLTQMY-GDYMTFPEKEMQEWYSH 261

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1216

A DNA sequence (GBSx1292) was identified in *S.agalactiae* <SEQ ID 3773> which encodes the amino acid sequence <SEQ ID 3774>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> May be a lipoprotein
 INTEGRAL Likelihood =-12.05 Transmembrane 554 - 570 (547 - 575)

----- Final Results -----
 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 SEQ ID 3774 (GBS182d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 8; MW 62kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1217

35 A DNA sequence (GBSx1293) was identified in *S.agalactiae* <SEQ ID 3775> which encodes the amino acid sequence <SEQ ID 3776>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4653(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1218

A DNA sequence (GBSx1294) was identified in *S.agalactiae* <SEQ ID 3777> which encodes the amino acid sequence <SEQ ID 3778>. This protein is predicted to be DOLICHYL-PHOSPHATE MANNOSE SYNTHASE RELATED PROTEIN. Analysis of this protein sequence reveals the following:

5 Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -2.92 Transmembrane 232 - 248 (231 - 248)
----- Final Results -----
10 bacterial membrane --- Certainty=0.2168 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9991> which encodes amino acid sequence <SEQ ID 9992> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA35924 GB:AF071085 putative glycosyl transferase [Enterococcus faecalis]
Identities = 118/240 (49%), Positives = 152/240 (63%), Gaps = 1/240 (0%)
20 Query: 14 KILLVIPAYNEBGSIAKTQVTIVDFKASRS-LPFELDVIYVINDGSTDGTPELLDRIGLH 72
K+LL+IPAYNER +I +T+ +I FK + ELDI+VINDGSTDGT ++L+ +N
Sbjct: 2 KLLIIPAYNEENILRTIASIETFKQEVTHPQHELDVIYVINDGSTDGTQILEVNVQINA 61
25 Query: 73 IDLVNVLGIGGCVQVTGYLYANRHHYDVAQVQDGDGKHDIRSIEDVVMFLINDEADFPVIGS 132
I LV NLGIGG VQTGY YA N YDVA QFDGSG HDI S+ ++ P+ F QS
Sbjct: 62 IHLVNLGIGGAVQVTGYKALENETDVAAQFDGSGKHDIKSLFILLEPLABGKCKESXGS 121
30 Query: 133 RFVDKXKHNQSTAMRRLGINLISAAIKLTGTHKIVDTTSGYRAANAALIAHLSCYFPVQ 192
RF+ ++QS MRR GI L+S G +Y T G RA N +IA+ + YP
Sbjct: 122 RFPGNDASTFQXKMRXGIRLLSFCCKXANGKTIYKVTGKRGANRKRVIATFAKRYPTN 181
Query: 193 YPEPESTARILKKGYRLKEVTANMFEREAGTSSISLESIFPMITVLTSIIIAAGFIKRED 252
YPEPES ++KK + + E NM ER G SSI +L S+ YM +V ++I+IA F+KB D
35 Sbjct: 182 YPEPESTIVELIKRPFVIVERFVNMMERLGGVSSIRALASVKIMLEVGSAIIAPFMKEGD 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3779> which encodes the amino acid sequence <SEQ ID 3780>. Analysis of this protein sequence reveals the following:

40 Possible site: 56
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.80 Transmembrane 211 - 227 (211 - 227)
----- Final Results -----
45 bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA35924 GB:AF071085 putative glycosyl transferase [Enterococcus faecalis]
Identities = 104/233 (44%), Positives = 134/233 (56%), Gaps = 9/233 (3%)
50 Query: 1 VKGLIIPAYNESSNIWVIRTIESDAFD-----FDYIIDDSTNDLTAICQKQGFH 53
+K L+IIPAYNE NI+ TI +IE+ + DY++I+D STD T I + N
55 Sbjct: 1 MKVLIIPAYNEENILRTIASIETFKQEVTHPQHELDVIYVINDGSTDGTQILEVNVQIN 60
Query: 54 VISLPIVLGIGGAVQVTGYRYAQRQSDVAVQVQDGDGHNPCYLEHNVVFLVQSSVNVIG 113
I L +NLGIGGAVQVTGY+YA YDVA Q DSGG H+ L ++E L + G
Sbjct: 61 AIHLVNLGIGGAVQVTGYKALENETDVAAQFDGSGKHDIKSLFILLEPLABGKCKESXGS 120

-1364-

Query: 114 SRPI--TKSGFOSSFARRIGIKYFTWLIALLTGGCITDATSGRLRIDRSlierFANHYPD 171
 SRPI FQS RR GI+ ++ G I T GR +R +I FA YP
 Sbjct: 121 SRPIFGXASFSQSKMRXGIRLLSCCKXMGXTIYXVTKGXRAGNKVIAPFAKRYPT 180

Query: 172 DYPEPEFVVDVLVSHFKVKELPVMNERQGGVSSISLTKSVYMIKVTAILV 224
 +YPEPE+V ++ F + E PV M ER QGVSSI SV YN+V AIL+
 Sbjct: 181 NYPEPESIVHLIKRPFVIVERFVMMERLGGVSSIRALASVKYMLEVGSAILI 233

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 105/231 (45%), Positives = 142/231 (61%), Gaps = 8/231 (3%)

Query: 14 KILLVIPAYNEBGSIAKTVQTVDFKASRSLPELDYIVINDGTDGTPELLDRGLNHI 73
 K L+IPAYNE +I T++TI S + DYI+I+D STD T + + G N I
 Sbjct: 2 KKLIIIPAYNESNTVNTIKTI-----ESDAPFDYIIIDD CSTDNTLAIQKQGFNVI 55

Query: 74 DLVQNLGIGGCVQGTGYLYANRNHYDVAQFDGGQHDIRSIEDVVPILNDEADFVIGSR 133
 L NLGIGG VQGTGY YR YDVAQ DGGQGH+ +E +V ++ + VIGSR
 Sbjct: 56 SLPLNLGIGGAVQGTGYRACQGYDVAQVDGQGNPCYLEKMEVLVQSSVNMVIGSR 115

Query: 134 FVDKKGKFNQFSQAMRLGILNLSAAIKLTTGHHKVDTTSGYRANAALIAYLSCHYVQY 193
 F + K + FQS+ RR+GI + I L TG K D TSG R + +LI + HYP Y
 Sbjct: 116 FITK--EGFQSSFARRIGIKYFTWLIALLTGKKTATSGRLRIDRSlierFANHYPDY 173

Query: 194 PEPESSTARILKGYRLKEVTANMFEREAGTSSISLKSIFMYTDLVLTSLII 244
 PEPES+ +L +++KE+ M ER+ G SSIS KS+YM V +I++
 Sbjct: 174 PEPETVVDVLVSHFKVKELPVMNERQGGVSSISLTSVYMIKVTAILV 224

A related GBS gene <SEQ ID 8751> and protein <SEQ ID 8752> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Cxend: 9
 MoG: Discrim Score: 0.29
 GvH: Signal Score (-7.5): -4.34
 Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -2.92 threshold: 0.0
 INTEGRAL Likelihood = -2.92 Transmembrane 222- 238(221 - 238)
 PERIPHERAL Likelihood = 4.40 4
 modified ALOM score: 1.08

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00548[340 - 1056 of 1359]
 GP[3608398][gb|AAC35924.1|]AP071085(2 - 241 of 241) putative glycosyl transferase
 [Enterococcus faecalis]
 %Match = 24.7
 %Identity = 49.2 %Similarity = 64.2
 Matches = 118 Mismatches = 85 Conservative Sub.s = 36

249 279 309 339 369 399 429 456
 L*QD*GGYGRMVIKINLSIKLCING*XQQIXIRIDRMKKILLVIPAYNEBGSIAKTVQTVDFKASRSLPELDYIV
 :||:||||| :| :|:| || : : |||:
 MKVLLIIPAYNEENTLRTIASISTFKGVTHFQSLDLYV
 10 20 30 40

486 516 546 576 606 636 666 696
 INDGSTDGTPELLDRGLNHIHDLVQNLGIGGCVQGTGYLYANRNHYDVAQFDGGQHDIRSIEDVVPILNDEADFVIGS
 ||||| :|: :| || ||||| |||| | |||| ||||| ||| : :| : ||

-1365-

```

INDGSDGTGKQILKFNQINAIHLVNLGLIGGAVQTGYKYALENEYDVAXQFDGDXHDXSLFLLEPLASGXCKXFXSGS
              60          70          80          90          100          110          120
5      726      756      786      816      846      876      906      936
      RFDKKHQNFSITAMRRGLINLSAAIKLTGHHKVVYDTTSYRAANAALIAYLSCHTYFVQYPEPESTARI LAKGKYRLKEV
      ||: ||| ||| ||| ||: ||| ||| ||: ||: ||| ||| ||: ||| ||: ||
      RFLPGNHASFSQSKMRRAIGLLGFCCKXAXGXTYYKVTGKXKRAKRVIAFAFKYPTMYPEPESIVHLIKKRFVIVRR
              140          150          160          170          180          190          200
10     966      996      1026      1056      1086      1116      1146      1176
      TANMFRRAGTSSISLKSIFMIDVLTSIIAGFPKEDDRV*V*HCKLCKLF*PLSYPT*L*EWLIRTHFLGLNLYLGY*
      ||| ||| ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
      FVMHMERLGGVSSIRALASVYKMLRFVGSAILIAPFMKEGD
              220          230          240
15

```

SEQ ID 8752 (GBS355) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 4; MW 27kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 7; MW 52kDa).

GBS355-GST was purified as shown in Figure 213 (lane 4) and in Figure 216 (lane 6).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1219

A DNA sequence (GBSx1295) was identified in *S. agalactiae* <SEQ ID 3781> which encodes the amino acid sequence <SEQ ID 3782>. Analysis of this protein sequence reveals the following:

```

25     Possible site: 19
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL      Likelihood = -1.91      Transmembrane 185 - 201 ( 185 - 201)

    ----- Final Results -----
30         bacterial membrane --- Certainty=0.1765 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35     >GP:BA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
        Identities = 181/315 (57%), Positives = 244/315 (77%), Gaps = 7/315 (2%)

    Query: 1   MKVNILMATYNGEKFLAQQIESIQKQTFKEWNLIRDOGSSDKTCDIIRNFTAKDGRIRP 60
              MKVNILMATYNG++P+AQQI+SIQKQTF+ WNLIRDOGSSD T II +F D+RIRP
40     Sbjct: 1   MKVNILMATYNGEFLAQQIQSIQKQTFENWNLIRDOGSSDKTFLIADFAKSDARIRP 60

    Query: 61   INENHNHNLGVISKFPITLVNYEVADPFYFQDDVDWLPEKLSVSLFAAKHKASDVPLLVY 120
              IN ++ N GVIK+F+TL+ YE AD+YFSDQDDVWLP+KL ++L + + + +FL+VY
45     Sbjct: 61   INADKRNFGVINKNFTLLKYEKADYFSDQDDVDWLPEKLSVSLFAAKHKASDVPLLVY 120

    Query: 121  TDLKVVNQELNLQDSMIRAQSHHANTLLPELTENTVGTGTMIMNHALAKKW-FTPNDI 179
              TDL VV+++L +L DSMI+ QSHHANT+LL ELTENTVGTGTM+NH LA++W + +D+
50     Sbjct: 121  TDLTVVDRDLQVLHDSMTKQSHHANTSLLSELTENTVGTGTMVNVHCLAKQWQCYCDL 180

    Query: 180  LMDWFLALLAASISGEIIYLDLPTQLYRHDNNVLGARTMDKRFK-ILRSGPKSIFTRYW 238
              +MHDW+LALLAAS+G++IYLD T+LYRQH++NVLGART KR K LR P + +YW
55     Sbjct: 181  LMDWFLALLAASISGLIYLDFTTELYRQHSNNVLGARTWSKRLKNLWR- -PHRLVKYVY 238

    Query: 239  KLHDSQKQASLIVDKYGDIMTANDELTKCFIKIDKQFPMTLRMLMKYGSRNQFQHQ 298
              L+ SQ+QAS +++ D+ AN +I+ ++ + Q F+ R++WL +YG++KN+ H
60     Sbjct: 239  WLVTSSQQQASHLEL--DLPAANK-ALIRAYVTLLDQSLNKRINKLQYGFARNRAPHT 295

    Query: 299  VVFWLIRLNTYYNKR 313

```

-1366-

VFK LI T + +R
 Sbjct: 296 VFVKTLITTFGYRR 310

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 817> which encodes the amino acid sequence <SEQ ID 818>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1980 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 178/314 (56%), Positives = 232/314 (73%), Gaps = 6/314 (1%)

Query: 1 NKNVILWATYNGEKFLAQQTESIQKOTFKENLLIRDDGSDKTCIIRNPTAKDSRIRF 60
 N + NIL+ + TNGE+ FLA+ QT+ SIQ+ QT +N LLIRDDGSD T DIIR F + D RI++
 Sbjct: 1 MNINILLSTYNGERFLAQQTESIQKOTFKENLLIRDDGSDTQDQDIRTFVKEDRIQW 60

20 Query: 61 INENHEINLGVIKSPFTLVNVEVADYFFPSDQDQVWLPEKLSVS-LEAIGHKASDVPLLV 119
 INE + NLGVIK+ F+ TL+ ++ AD YFFPSDQDQ+ WL KL V+ LEA KH+ + PLLV
 Sbjct: 61 INEQGTENLGVIKHFTLLKHQKADYFFPSDQDQIMLDNKLAVTLLEAKHMT-APLLV 119

25 Query: 120 YTDLKVNVQELNLIQDSMIRAGSHHANTLLPELTENTVTGGTMMINIALAEKWTFPNDI 179
 YTDLKVQ Q L + DSMI+ QS HANT+ LL ELTENTVTGGTMMI HALAE+ M T + +
 Sbjct: 120 YTDLKVVTQHLAVCHDSMIKTQSGHANTSLQELTENTVTGGTMMITHALAEWTTCDGL 179

30 Query: 180 LMDHWFALLAASLGIIYLDLPTQLYRQHNNVLGARTMDKRPFKLREGPKSIFTRYWK 239
 LMDH+ LALLA++ +G++ YLD+ PT+ LYRQHD NVLGART KR K P + +YW
 Sbjct: 180 LMDHWFALLAASLGILVYLDIPTLYRQHNDANVLGARTWSKRMNMT-PHLLVKNYWM 238

Query: 240 LIHDSQKQASLIVDKYGDIMTANULELIKCFIKIDQPFMTLRNLWKYGYSEKQFKHQV 299
 LI SQKQA L+D + ND EL+ + + PF RL L +YQ+ KM+ H
 35 Sbjct: 239 LITSSQKQQLLLDL---PLKPNDELHVTAYVSLDNPPTKRLATLKYGFPIKIRIFHTF 295

Query: 300 VFKNLAINVYNNR 313
 +F+ L+ T + +R
 40 Sbjct: 296 IFRSLVLTFLGYRR 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1220

45 A DNA sequence (GBSx1296) was identified in *S. agalactiae* <SEQ ID 3783> which encodes the amino acid sequence <SEQ ID 3784>. This protein is predicted to be rgpAc. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.1881 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9993> which encodes amino acid sequence <SEQ ID 9994> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1367-

>GP:BA32089 GB:AB010970 xgphc [Streptococcus mutans]
Identities = 234/362 (64%), Positives = 264/362 (77%)

Query: 33 VSELINHQKSPDIKHYVACLSDKEHHMFNADADCFITNPQLGPARVIAYDMAINYA 92
+ EL+ ++S + YVACLSD+ H HF + DCFTI P+LGPARVIAYD+MAI YA
Sbjct: 1 MEELVKYKQSQQLTYEVACLSETDQKHFTYLVGDCFTIKAPKLGPARVIAYDMAIRYA 60

Query: 93 LDINVKTHDLKEPIFYILGNTIGAPFIWHFANKIHRVGGLLVNPDLGLENKRSKSRPTORY 152
L L+K +K PIFYLGNITGA+ FA KI ++GG Y+NPDLGLEN+RSKSRP Q Y
Sbjct: 61 LKLIKQDKIKHPFIYILGNTIGAPMGPFAKIKRIGRFYINPDGLKRSKSRPQVAY 120

Query: 153 LKVAEKMTQADLISDNIGIENVIQSTYSNVKRTFIAYGTINSRKLSSDPRVQLF 212
LKVAEKMTK ADL+ISDN GIK YI+ Y KT FIAYGT+++ L +D +YK +
Sbjct: 121 LKVAEKMTYKADLVISDNIGIENVIQSTYSNVKRTFIAYGTDLSPGLLNDKVKDFY 180

Query: 213 KKNWIKSKGYLIVGRFVPPNNYETAIREFMASDTKROLVILCNHQNPFYFKLSLKTNL 272
KKN IK KGYLIVGRFVPPNNYETAIREFM S ++RDLVILCN+ N YF L KT
Sbjct: 181 KKNWIKSKGYLIVGRFVPPNNYETAIREFMSTSSERDLVILCNHQNPFYFKLSLKTNL 240

Query: 273 QQDKRVKFGVGLYEKDLIDYVROQAFYIHGHVGGTINPGLLEALANTDLNLDVDFNK 332
+DKR+KFGVT+Y++ LL Y+R+QAFYIHGHVGGTINPGLLEALA+DLNLDL -FN
Sbjct: 241 DKDKRIKFGVTYDRFLTYIRQAFYIHGHVGGTINPGLLEALANTDLNLDVDFNK 300

Query: 333 SVAGLSSFYNAKGEGLAKLINDSQQDLSTYGDRAKIIQENYTWKIVSEVEDFLN 392
+VA ++ YW + G LA+LIN D+Q++ + YG RAK II YTW+KIVSEVEDFLN
Sbjct: 301 TVALDAARYWTQNGSLAQINQDKENFAEYQRAKEITVNYTWKIVSEVEDFLN 360

Query: 393 ES 394
ES
Sbjct: 361 ES 362

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3785> which encodes the amino acid sequence <SEQ ID 3786>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.38 Transmembrane 95 - 111 (95 - 111)

----- Final Results -----
bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 250/383 (65%), Positives = 307/383 (79%)

Query: 11 MQDVFIISGRGLPARYGSPFTFVSELINHQKSPDIKHYVACLSDKEHHMFNADADCFIT 70
MQDVFIISGRGLP+YGGSPFTFV ELI+HQ S +I+YHVACLSD +H HF++ ADCF
Sbjct: 1 MQDVFIISGRGLPARYGSPFTFVSELINHQKSPDIKHYVACLSDTHKIVLIPYQKADCFY 60

Query: 71 INPQLGPARVIAYDMAINVALDLVKTDLKEPIFYILGNTIGAPFIWHFANKIHRVGGLL 130
+NPP+LGPARVIAYD+MAI YAL H ++ PIFV+LGNTIGAPFI +IH GG
Sbjct: 61 LNPPLGPARVIAYDMAITVALSYSDHQIQNPFIYILGNTIGAPFIADPFVQLHRRGR 120

Query: 131 LVYNPDGLKRSKSRPTORYLKVAEKMTQADLISDNIGIENVIQSTYSNVKRTFI 190
++NPDLGLENKRSKSRP Q YL++EK MT+ ADL+ISDNIGI+ Y++ Y KT FI
Sbjct: 121 FFNPDLGLENKRSKSRPQVAYLKFSEKMTQADLVISDNIGIIRYLVKQVPSKTCFI 180

Query: 191 AYGTEINSRKLSSDPRVQLFKKNWIKSKGYLIVGRFVPPNNYETAIREFMASDTKRD 250
AYGT+ ++L+D +Y+ F+ ++I+ K YYL+GRFVPPNNYETAIREFMAS TKRD
Sbjct: 181 AYGTTQPSRLATADSKVRAYPQTDFIREKDYLLIGRFVPPNNYETAIREFMAS TKRD 240

Query: 251 LVILCNHQNPFYFKLSLKTNLQDKRVKFGVGLYEKDLIDYVROQAFYIHGHVGGTIN 310
LVILCNH+ N YF++L +T +D R+KFGVGLYK+LL Y+R+Q+AYIHGHVGGTIN
Sbjct: 241 LVILCNHGNAYFKQLLAETECNDPRIFKFGVGLYDKRLIYAIRQAFYIHGHVGGTIN 300

-1368-

Query: 311 PGLLEALANTDLNLVLVDVDFNKSVAIGLSSPYWAKGGDIKLLINDSDQQQLSTYGDRAK 370
 PGLLEALA-T+LNLVL VDFN+SWA ++ YW K+G LA+LIN D D G AK
 Sbjct: 301 PGLLEALANTDLNLVLVDVDFNQSVAKSAALYWTCKGQQLAELINQVDAGFSDHLKKEAK 360

5 Query: 371 AIIQEHYTWKIVVEEYDLPLNE 393
 AIIQE+YTW+KIV EYE LPLNE
 Sbjct: 361 AIIQEHYTWKIVGYEALPLNE 383

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1221

A DNA sequence (GBSx1297) was identified in *S.agalactiae* <SEQ ID 3787> which encodes the amino acid sequence <SEQ ID 3788>. This protein is predicted to be dTDP-L-rhamnose synthase. Analysis of this protein sequence reveals the following:

15 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1059 (Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAD10164 GB:AF026471 Cps20 [Streptococcus pneumoniae]
 Identities = 258/283 (91%), Positives = 274/283 (96%)

Query: 1 MILITGANGQLGSELRLHLLDERTQBYVAVDVAEMDITNAEMVDKVFEEVKPFLVYHCAR 60
 MILITGANGQLG+ELR+LLDER +EYVAVDVAEMDIT+AEV+KVFEEVKP+LVYHCAR
 Sbjct: 1 MILITGANGQLGTELRYLLDERNEEYVAVDVAEMDITDAEMVEKVFEEVKPFLVYHCAR 60

30 Query: 61 TAVDAAEDEGKELDFAINVTGTENNAKAAKHDAITLVYSTDYVFDGKFPVQGENVEDDL 120
 TAVDAAEDEGKELDFAINVTGT+NVAKA+ KH ATLVLVYSTDYVFDG+KPVQGENVEDD
 Sbjct: 61 TAVDAAEDEGKELDFAINVTGTGNNAKASEKHGATLVYSTDYVFDGKFPVQGENVEDDR 120

35 Query: 121 PDPKTEYGRTRKMGEEELVEKYTSKFYTI RTAMVFGNYGKGFVFTMQNLAKTHKTLTVVND 180
 PDP+TEYGRTRKMGEEELVEK+ S FY IRTAMVFGNYGKGFVFTMQNLAKTHKTLTVVND
 Sbjct: 121 PDPQTEYGRTRKMGEEELVEKHVSNFYLI RTAMVFGNYGKGFVFTMQNLAKTHKTLTVVND 180

40 Query: 181 QHGRPTWTRTLAEFNTYLAENKQKDFGYHLSNDAKSDTTWYDFAVEILKOTDVEVKPVD 240
 Q+GRPTWTRTLAEFNTYLAEN+K+FGYHLSNDA EDTWYDFAVEILKOTDVEVKPVD
 Sbjct: 181 QHGRPTWTRTLAEFNTYLAENKKEFGYHLSNDAEDTTWYDFAVEILKOTDVEVKPVD 240

Query: 241 SQPFAKARRPLNSTMSLEKAKATGFPVIPTWQDALKEPFYQEVK 263
 SQPFAKARRPLNSTMSL KAKATGFPVIPTWQDAL+EPYQEV+
 45 Sbjct: 241 SQPFAKARRPLNSTMSLAKAKATGFPVIPTWQDALKEPFYQEVK 263

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3789> which encodes the amino acid sequence <SEQ ID 3790>. Analysis of this protein sequence reveals the following:

50 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0618 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/284 (79%), Positives = 248/284 (86%)

-1369-

Query: 1 MILITGANGQLGSELEHLLDERTQYVAVDVAEMDITNAEMVDKVPFESVPSLVHCAAY 60
 Sbjct: 21 MILITGANGQLG+ELR+LLDER +YVAVDVAEMDITN + V+ VF +VKP+LVHCAAY 80

Query: 61 TAVDAAEDBEGKELDFAINVTGTENAKAAKHDAITLVISTDYVFDGKPKVQGEWVDL 120
 Sbjct: 81 TAVDAAEDBEGK L+ AINVTG+EN+AKA K+ ATLVIISTDYVFDG KPVQGEW D 140

Query: 121 PDKTEYGRTRKMGELVEKYTSKFYTIRTAWVFCNGYGNFVFTMQNLAKTHILTVND 180
 Sbjct: 141 PDKTEYGRTRK+GE VE+Y FY IRTAWVFCNGYGNFVFTM+ LA+ H LTVND 200

Query: 181 QHGRPTWITRLAEFMVTLAENKQDNGYHLSNDAKEDTWTYDFAVRLIKDTDVEVKPVD 240
 Sbjct: 201 QHGRPTWITRLAEFMVTLAENKQDNGYHLSNDAKEDTWTYDFAVRLIKDTDVEVKPVD 260

Query: 241 SQPPAKAKRPLNSTMSLEKAKATGFVIFTWQDALKFYEQVKK 284
 S FPAKAKRPLNSTM+L+KAKATGFVIFTWQ+ALK FY+Q +KK
 Sbjct: 261 SAPPKAKRPLNSTM+L+KAKATGFVIFTWQ+ALKFY+Q+KK 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1222

- 25 A DNA sequence (GBSx1298) was identified in *S. agalactiae* <SEQ ID 3791> which encodes the amino acid sequence <SEQ ID 3792>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2554 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA21508 GB:AB000631 unnamed protein product [Streptococcus mutans]
 Identities = 92/108 (85%), Positives = 100/108 (92%)

40 Query: 5 KQYSEEVGKIKDRILEALEMVIDPELGIDIVNLGLIYIRFEDNGRTEIDMTLTMGCP 64
 K Y+ EE+ KIKDRILEALEMVIDPELGIDIVNLGLIY+IRFED+GRTEIDMTLTMGCP
 Sbjct: 4 KNTVEEIAKIKDRILEALEMVIDPELGIDIVNLGLIYDIRFEDSGRTEIDMTLTMGCP 63

Query: 65 LADLLTDQIHQDVKNTKPEVTEVVKLVNYPANGVDKMSRYARIALGIR 112
 LADLLTDQIHQD +K VPEV +VKLWN PAK+VDKMSRYARIALGIR
 45 Sbjct: 64 LADLLTDQIHQDVKPEVLDIVKLWNSPAKTVDKMSRYARIALGIR 111

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3793> which encodes the amino acid sequence <SEQ ID 3794>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2818 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 90/112 (80%), Positives = 102/112 (90%)

-1370-

Query: 1 MSEVKQKSEBEEVGKIKDRILEALEWVIDPELGLIDIVNLGLIYEIRFEDNGRTEDMTLTT 60
 NS+ +Y++++V IK+RILEALE VIDPELGLID+VNLGLIYEIRF DNG TRIDMTLTT
 Sbjct: 1 MSDTFKYTDQVIAIKNRILEALETVIDPELGLIDVNLGLIYEIRFNDNGYTEDMTLTT 60

5 Query: 61 MGCPALDLITDIDHDVMKTVPEVETEVKLWVYPAWSVDKMSRYARIALGIR 112
 MGCPALDLITD IDH +VPEVT+TEVKLWVYPAW+VDKMSRYARIALGIR
 Sbjct: 61 MGCPALDLITDYIHDLQVPEVTKTEVKLWVYPAW+VDKMSRYARIALGIR 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1223

A DNA sequence (GBSx1299) was identified in *S. agalactiae* <SEQ ID 3795> which encodes the amino acid sequence <SEQ ID 3796>. This protein is predicted to be RNA polymerase sigma factor, sigma-70 family (rpoD). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3157 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein is similar to the sigma-42 protein from *S. mutans*:

>GF:BAA21507 GB:AB000631 sigma 42 protein [Streptococcus mutans]
 Identities = 345/367 (94%), Positives = 358/367 (97%)

Query: 14 EKKQNTTFNVQVADFIRNHKKQQTAAIDDEVTEKLVIFFVLDAQDIDLLERLTGGISIT 73
 +KK +TFNVQVADFIRNHKK+G A+DDEVTEKLVIFF L+A+QIDLLERLTGGISIT
 Sbjct: 5 KKKTSSTTFNVQVADFIRNHKKEGVAVDDEVTEKLVIFFLEABQDIDLLERLTGGISIT 64

30 Query: 74 DKGNPSTKYVVEGPKPEELTDELLIGNSAKVNDPVRMYLKEIGVVPLLTNEEEKELAV 133
 D+EGNPSTKY VS KPEELTDEEL+GNSAKVNDPVRMYLKEIGVVPLLTNEEEKELAV
 Sbjct: 65 DREGNSTKYAVEEIKPEELTDELLIGNSAKVNDPVRMYLKEIGVVPLLTNEEEKELAI 124

35 Query: 134 AVAEGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIOBGNMGLMKAVDKFDYSEKGF 193
 AV GDL AKQRLAEANLRLVVSIAKRYVGRGMQFLDLIOBGNMGLMKAVDKFDYSEKGF
 Sbjct: 125 AVENGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIOBGNMGLMKAVDKFDYSEKGF 184

40 Query: 194 FSTYATWIRQAITRAIDQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER 253
 FSTYATWIRQAITRAIDQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER
 Sbjct: 185 FSTYATWIRQAITRAIDQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER 244

45 Query: 254 MDMTFDKVRKELAKIAQEPVSLETPIGREDDSHLGDFIDEVNIENFVDYTRVVLRLQDLE 313
 MDMTFDKVRKELAKIAQEPVSLETPIGREDDSHLGDFIDEVNIENFVDYTRVVLRLQDLE
 Sbjct: 245 MDMTFDKVRKELAKIAQEPVSLETPIGREDDSHLGDFIDEVNIENFVDYTRVVLRLQDLE 304

50 Query: 314 VLDTLTDRENVILRLPGLDDGKMTLEDVGKVFVNTVRERIRQIEAKALRKLHPSRSKQ 373
 VLDTLTDRENVILRLPGLDDGKMTLEDVGKVF+VTRERIRQIEAKALRKLHPSRSKQ
 Sbjct: 305 VLDTLTDRENVILRLPGLDDGKMTLEDVGKVFVNTVRERIRQIEAKALRKLHPSRSKQ 364

Query: 374 LKDFMED 380
 L+DF+ED
 Sbjct: 365 LRDFVED 371

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3797> which encodes the amino acid sequence <SEQ ID 3798>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

-1371-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1788(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 351/369 (95%), Positives = 364/369 (98%)

Query: 12 MAEKGNITFNQVADFRNHKKQGTADDSVTEKLVIPFVLADQIDDLERLTDGGIS 71
 M ++K TFPNVQVA+FTIR+HKK+GTAIDD+VTEKLVI PF LADQIDDLERLTDGGIS
 Sbjct: 1 MTKQKEITFPNVQVAEFTIRHKKEGTAIDDDVTEKLVI PFALDADQIDDLERLTDGGIS 60

Query: 72 ITDKGNPSTICVVBGPKPEELTDEELIGSNSAKVNDPVRMYLKBIGVVPVLLTINSEKEL 131
 ITDKGNPS+KY+VE PKPEELTDEELIGSNSAKVNDPVRMYLKBIGVVPVLLT+EEKEL
 Sbjct: 61 ITDKGNPSSKYIVBEPKPEELTDEELIGSNSAKVNDPVRMYLKBIGVVPVLLTSEEEKEL 120

Query: 132 AVAVAGDLMAKQRLAEANRLRVVSIARKYVGRGMQFLDLIQSGNMGIMKAVDKFDYSKG 191
 AVAVA+GDLMAKQRLAEANRLRVVSIARKYVGRGMQFLDLIQSGNMGIMKAVDKFDYSKG
 Sbjct: 121 AVAVAGDLMAKQRLAEANRLRVVSIARKYVGRGMQFLDLIQSGNMGIMKAVDKFDYSKG 180

Query: 192 FKFSYATWIRQAITRAIDQARTIRIFVHMVETINKLVREQRNLLQELGQDPTPEQIA 251
 FKFSYATWIRQAITRAIDQARTIRIFVHMVETINKLVREQRNLLQELGQDPTPEQIA
 Sbjct: 181 FKFSYATWIRQAITRAIDQARTIRIFVHMVETINKLVREQRNLLQELGQDPTPEQIA 240

Query: 252 ERMDMTFDKVRKILKIAQSFVSLPTPIGEEDDSHLGDFIDEVIENFVDYTRVVLREQL 311
 ERM+MTFDKVRKILKIAQSFVSLPTPIGEEDDSHLGDFIDEVIENFVDYTRVVLREQL
 Sbjct: 241 ERMDMTFDKVRKILKIAQSFVSLPTPIGEEDDSHLGDFIDEVIENFVDYTRVVLREQL 300

Query: 312 DEVLDTLTDRREENVLRRLFGLLDGGKMTLEDVGKVFNVTRERIRQISAKALRKLHPSPRS 371
 DEVLDTLTDRREENVLRRLFGLLDGGKMTLEDVGKVFNVTRERIRQISAKALRKLHPSPRS
 Sbjct: 301 DEVLDTLTDRREENVLRRLFGLLDGGKMTLEDVGKVFNVTRERIRQISAKALRKLHPSPRS 360

Query: 372 KQLRDFIED 380
 KQL+DF+ED
 Sbjct: 361 KQLRDFIED 369

Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 1224

40 A DNA sequence (GBSx1300) was identified in *S. agalactiae* <SEQ ID 3799> which encodes the amino acid sequence <SEQ ID 3800>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2853(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1372-

Example 1225

A DNA sequence (GBSx1301) was identified in *S.agalactiae* <SEQ ID 3801> which encodes the amino acid sequence <SEQ ID 3802>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2196 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA03516 GB:D14690 DNA primase [Lactococcus lactis]
  Identities = 206/398 (51%), Positives = 294/398 (73%), Gaps = 6/398 (1%)

Query: 37 LAIDKEKISEIKNSVNIQVIGEVVGLTKTGRNHLGLCPFHKEKTPSFNVIEDQFFHCF 96
      +++D E ++++K+ VNI D+I + V L++TG+N+GLCPFH EKTSPFNV ++ F+HCF
Sbjct: 2 VSLDTEVNDLKSQVNIADLSIQVLSRTGKNIYIGLCPFHKEKTPSFNVNAEKGFFHCF 61

Query: 97 GQSRSGDVFKFVEDYQHISFLDSVQVLAERSGIPLDTNFKQVPPKKPAKNGSILLDIRVA 156
      GQSRSGD +F+++Y + F+D+V+ Lh+ +G+ L N ++K N L +I+ A
Sbjct: 62 GQSRSGDAIEFLKEYNQGVFVDVKELADFAGVTL--NISDREERINPNAPLFEINQQA 119

Query: 157 SGFYHAYLMTINDGERARQYLAERGVTDLIKHFGQIGLSPGGQDFLYRLAKEFEDEKTL 216
      + Y+ Lm+T GERAR+YL ERG+T+D+IK F IGL+P DF+++ L+ +FDE+
Sbjct: 120 ARLYNILLMSTELGERARKYLEERGITDDVIKRFNIGLAPEENDFIKILSNKPFDEEIMA 179

Query: 217 SSGLFNYSNSNQFYDSFNNRIMFPLINDIGEVVIAFSGRVVTQEDIDRQKQAYKNSPATP 276
      SGLF++S +N+ +D+P NRIMFP+TN+ G+ I FSGR W QE+ D K AKY N+ AT
Sbjct: 180 KSGLFHFS--NNKVFDPATNRIMFPITNEGGTIGFSGRW--QEDDSK--AKYINTS 235

Query: 277 IFNKSVELYHLDKARAVINKAHEVYLMBSGFMVDVIAAYAGIENNVASMGITALINNHVRL 336
      IF+KSYEL+LDKA+ I+K HEVYLMBSGFMVDVIA+Y+AGI NVVASMGTALT +HVR L
Sbjct: 236 IFDKSYELWNLDKAKPTISKQHEVYLMBSGFMVDVIAKYAGINNVASMGITALTEGVRL 295

Query: 337 KRFTKKVVLTYDGDRAQNAIDKSLELLSMTVDIVIRIPNMQDPEFLQANSABDFQQLL 396
      K+ KK VL YDGD AGQNAI K+++L+ + V IV++P +DPDE+ + + L+
Sbjct: 296 KQMAKFKVLVYDGDCAQNAIYKAILDIGESAVQIVKVPBGDLDPDEYSKNYGLKLSALM 355

Query: 397 ENGRISNTFPIHYLKPDNTDLQSEIAYVEKIARKLIA 434
      E GRI EF I YL+PEN NLQ+++ ++E+I+ +IA
Sbjct: 356 ETGRIQPIEFILDYLRPENLANLQTLQDFIEQTSFMTA 393

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3803> which encodes the amino acid sequence <SEQ ID 3804>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3532 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 378/604 (62%), Positives = 477/604 (78%), Gaps = 2/604 (0%)

Query: 28 MOYFGGHDLAIDKIKISEIKNSVNIQVIGEVVGLTKTGRNHLGLCPFHKEKTPSFNVI 87
      MG+ GG DLAIKDS IS++KNSVNIQVIGEVV L++G++LGLCPFHKEKTPSFNV+
Sbjct: 1 MQLFGGDDLAIDKIKSQVNSVNIQVIGEVVGLSRSGRHYLGLCPFHKEKTPSFNVV 60

Query: 88 EDRQFFHCFCGQSRSGDVFKFVEDYQHISFLDSVQVLAERSGIPLDTNFKQV--PKKPKA 145

```

-1373-

EDRQFFHCPCGG+SGDVFKP+E+Y+ + FL+SVQ+!A++G+ L+ V +
 Sb|ct: 61 EDRQFFHCPCGGSGDVFKPIEYRQVPFLESVQITADKIGMSINIPPSQAVLASQKHHP 120

Query: 146 NQSLDIDHRVASGPFYHAYLMTNDGGERARQVIAERGVETDLIKHPQIGLSGQGDPIYRR 205
 N+L+ +H A+ FYHA LMTT G+ AR+YL +RG+ + LI+HF IGL+P D+LY+
 5 Sb|ct: 121 NHALMTLHEDAAKFYHAYLMTTTTIGQARKYLYQRGLDQGLIEHPHIGLAPDSRDYLOA 180

Query: 206 LARGPDEKTIASGGLFNYGSENSQFYDGFNRIMFPLINDIGEVIAFSGRWVWTQEDIRK 265
 L+K+++E L++SGLF+ G+ SN YD+F NRIMFPL++D G +IAFSGR+WT D++++
 10 Sb|ct: 181 LSKKYREGQLVASGLFHLSDQSNITTYDAFNRIMFPLSDDRGHITAFSGRIWTAADMEKR 240

Query: 266 QAKYKNSRATPIFNKSYELVHLDKARAVINKAHEVYLMGPMVDVIAAYRAGIENVVASMG 325
 QAKYKNSR T +FNKSYELVHLAKAR VI K HV+LMGPMVDVIAAYR+G EN VASMG
 15 Sb|ct: 241 QAKYKNSRGTVLFNKSVELYHLDKARPIVIACTHEVFLMGPMVDVIAAYRSGYENAVASMG 300

Query: 326 TALTNHGVHRLGKPTKKVVLTYDGRAGQNAIDKSLSEDMTVDIVIRIPNKMDDPEFLQ 385
 TALT EHV HLK+ TKKVV L YDGD AGQ+AI KSELL D V+IVIRIPNKMDDPEF+Q
 20 Sb|ct: 301 TALTOEHVHRLGKPTKKVVLTYDGDAGQHAIAKSELLKDPVVEIVIRIPNKMDDPEFVQ 360

Query: 386 ANSADPFQKLENGRISNTEPYTHYLKPDENTHLOGEIAYVEKIAKLIAKSPSITAQNSY 445
 +S E F LL+ RIS+ EF+I YLKP N DNLQ+S I YVEK+A LIA+SPSITAQ+SY
 25 Sb|ct: 361 RHPSEAFADLLKQSRISSEVFYDILKPTNVNLOSQIVVEKMAPLIAQSPSITAQHSY 420

Query: 446 ITKVALLPDPDFYQVQESVNNERLHRSQQQASSSVQTSATVQLPOTGKLSAITKTEMQ 505
 I K+A+LLP+FDYFQVQESVN R+ R+ Q + S V LP L+AI KTE
 30 Sb|ct: 421 INKIADLLPDPDFYQVQESVNAIRIQQRKHQQIAQAVSNLVTLPMPKSLAITAKTESH 480

Query: 506 LPHRLNLHPYLLNEFRNDFYFCTTBIOVLVELLKGSEITSYLSQSDKVMRTYII 565
 L HRL+H YLINEFR+RD+PYFDT+ +++LY+ LK+ G ITS YDLS+ S++VNR TY +
 35 Sb|ct: 481 LMRHLHEDYLLNEFRHDDFYFDTSTLELLYQLKQQCHITSYLSQSEMSEHVNRYINV 540

Query: 566 LEEQLPVEVSIGIEIAVEKARDRLKRDRLKQSLIQSSNQGDREGALAALENIAQK 625
 LEE LP EV++GEI+ + R+LL ERDL KQ + +R+SSN+GD + AL LE+ IAQK
 40 Sb|ct: 541 LEENLPKEVALGEIDDLISKRAKLLAERDLKQGGKVRSSNKDREGAALVEHPHIAQK 600

Query: 626 RNME 629
 R ME
 Sb|ct: 601 RNME 604

- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1226

A DNA sequence (GBSx1302) was identified in *S. agalactiae* <SEQ ID 3805> which encodes the amino acid sequence <SEQ ID 3806>. Analysis of this protein sequence reveals the following:

45 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.05 Transmembrane 41 - 57 (34 - 58)
 INTEGRAL Likelihood = -5.79 Transmembrane 93 - 109 (90 - 112)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 55 A related GBS nucleic acid sequence <SEQ ID 9995> which encodes amino acid sequence <SEQ ID 9996> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC38560 GB:AF029731 large conductance mechanosensitive channel
 [Staphylococcus aureus]
 60 Identities = 64/126 (50%), Positives = 83/126 (65%), Gaps = 8/126 (6%)

-1374-

Query: 23 MIKELKEFLPKGNVLDLAVAVILGAAPNAITSLVKDVITPLILNPLVKAAGVSNIA-QL 81
 M+KE KEF KGNVLDLA+AV++GAPN II+SLV++I ELI K G + A +
 5 Sbjct: 1 MIKEPKFALKGNVLDLAI+VVMGAAFNKIISSLVNTIMELI-----GKIPGSDFAEW 56

Query: 82 SWNGVAYGNFLSAVINFLVGTTLFFPIVKAANKVMAKPAHEEIIIVETPQQLAEIR 141
 S+ G+ YG F+ +VI+P+I+ LF VK AM +M K+ AEE E V LL EIR
 Sbjct: 57 SPWGIKYGLPTQSVDFIILAPALFVVKIANTMKKKEAEE---ERVVEENVLLTEIR 113

10 Query: 142 DLLANK 147
 DLL K
 Sbjct: 114 DLLREK 119

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3807> which encodes the amino acid
 15 sequence <SEQ ID 3808>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -5.95 Transmembrane 71 - 87 (67 - 90)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.3378 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:
 >GF:CAB15653 GB:Z99122 similar to large conductance mechanosensitive
 channel protein [Bacillus subtilis]
 Identities = 61/126 (48%), Positives = 77/126 (60%), Gaps = 7/126 (5%)

30 Query: 1 MVKELKAPLFRGNIILAVAVILGGAFAIVTSFVNDIITPLILNPLKAAKNVENITQLS 60
 M E KAF RGNII+LA+ V+IGGAPG IVTS VNDII PL+ L + ++
 Sbjct: 1 MWNEKAPFAMRGNIVDLAIGVVGGAFAIVTSFVNDIITPLILNPLVGLLGLDPSGLSFTF 59

35 Query: 61 WNG-VKYGSP+ ++NPLII S+P V++ KKE E A QRELL EIR 114
 + VKYGSP+ ++NPLII S+P V++ KKE E A QRELL EIR
 Sbjct: 60 GDAVVKYGSPFQITVNFIIISPIPIVIRTNGLRKKKEAEEAEDAQAQRELL EIR 119

Query: 115 DLLAQK 120
 DLL Q+

40 Sbjct: 120 DLLKQ 125

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/125 (68%), Positives = 99/125 (78%), Gaps = 5/125 (4%)

45 Query: 23 MIKELKEFLPKGNVLDLAVAVILGAAPNAITSLVKDVITPLILNPLVKAAGVSNIAQLS 82
 M+KELK FLF+GN++LAVAVI+G AF AI+TS V D+ITPLILNP LKAA V NI QLS
 Sbjct: 1 MVKELKAPLFRGNIILAVAVILGGAFAIVTSFVNDIITPLILNPLKAAKNVENITQLS 60

50 Query: 83 WNGVAYGNFLSAVINFLVGTTLFFPIVKAANKVMAKPAHEEIIIVETPQQLAEIRD 142
 WNGV YG+FL AVINFLI+GT+LFF+VKAA K M KK E PTQE+LL EIRD
 Sbjct: 61 WNGVYKGSPLSAVINFLIGTSLFFVVKAAKAMPKKKK-----EAAAPTQRELLTEIRD 115

Query: 143 DLLANK 147
 LLA K

55 Sbjct: 116 LLAQK 120

A related GBS gene <SEQ ID 8753> and protein <SEQ ID 8754> were also identified. Analysis of this
 protein sequence reveals the following:

60 Lipop: Possible site: -1 Crend: 10
 SRCFLG: 0
 MCG: Length of UR: 4
 Peak Value of UR: 2.96

-1375-

```

Net Charge of CR: 1.39
McG: Discrim Score: 4.39
GVH: Signal Score (-7.5): -1.79
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 26
ALOM program count: 1 value: -5.79 threshold: 0.0
      INTEGRAL Likelihood = -5.79 Transmembrane 71 - 87 ( 68 - 90)
      PERIPHERAL Likelihood = 1.06 28
modified ALOM score: 1.66
icml HYPID: 7 CFP: 0.331

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00541(367 - 741 of 1041)
SP|O68285|MSC_L_STRAU(1 - 119 of 120) LARGE-CONDUCTANCE MECHANOSENSITIVE CHANNEL.
GP|3135292|gb|JHC38560.1|JAF029731 large conductance mechanosensitive channel.
{Staphylococcus aureus}
%Match = 14.9
%Identity = 53.3 %Similarity = 70.5
Matches = 65 Mismatches = 31 Conservative Sub.s = 21

177 207 237 267 297 327 357 387
QVMTSREITHSYPTFYDIIFSLOCKFFKLFQSGLLH*FNKIYR*FETYYLDPSKEICYNRELNNIKELVHMNIKLE
||||||
MKKSFKE

417 447 477 507 537 561 591 621
FLFKGNVLDLAVAVILGAAFNAITSLVKDVITPLIINPLVGAGVSNIAQLSWN--GVAYGNPLSAVINFLVGTLEFF
|:|||||||::|::| ||:|::|::| | | : |:|:|::|::| |
FALKGNVLDLALAIVMGAAFNKIISSLVENIMPLI---GKIFGSVDPAK-EWSFWGIKYLETQSVDFIIIAFALEFI
20 30 40 50 60 70 80

651 681 711 741 771 801 831 861
IVVAANKVMAKPKFEREEIEVVVEPTQGLAKEIRDLLANK**KTRITEFFPY*LIVIITYEKTAQG*TVFSYSI*LEFFTFA
||:||:|::|::| | | |||||
FVKIANILMKKEEAERE--AVVE-ENVVLEIRIDRLREKK
100 110 120

```

SEQ ID 8754 (GBS354) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 3: MW 17kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 **Example 1227**

A DNA sequence (GBSx1303) was identified in *S. agalactiae* <SEQ ID 3809> which encodes the amino acid sequence <SEQ ID 3810>. This protein is predicted to be 30S ribosomal protein S21-related protein. Analysis of this protein sequence reveals the following:

```

Possible site: 29
55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.6479(Affirmative) < success
      bacterial membrane --- Certainty=0.0000(Not Clear) < success

```

-1376-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9391> which encodes amino acid sequence <SEQ ID 9392> was also identified. A related GBS nucleic acid sequence <SEQ ID 10799> which encodes amino acid sequence <SEQ ID 10800> was also identified.

The protein is similar to the 30S ribosomal protein S21 from *Listeria monocytogenes*:

>GF:BA02793 GB:AB023064 30S ribosomal protein S21 [*Listeria monocytogenes*]
Identities = 30/34 (88%), Positives = 34/34 (99%)

Query: 1 MTKAGTLQESRKREFYEKPSVGRKKKSEAARKRK 34
++K+GTLQESRKREFYEKPSVGRKK+KSEAARKRK
Sbjct: 23 VSKSGTLQESRKREFYEKPSVGRKKKSEAARKRK 56

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3811> which encodes the amino acid sequence <SEQ ID 3812>. Analysis of this protein sequence reveals the following:

Possible site: 38
>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4815(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 35/36 (97%), Positives = 36/36 (99%)

Query: 1 MTKAGTLQESRKREFYEKPSVGRKKKSEAARKRK 36
+TKAGTLQESRKREFYEKPSVGRKKKSEAARKRK
Sbjct: 35 VTKAGTLQESRKREFYEKPSVGRKKKSEAARKRK 70

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1228

A DNA sequence (GBSx1304) was identified in *S.agalactiae* <SEQ ID 3813> which encodes the amino acid sequence <SEQ ID 3814>. Analysis of this protein sequence reveals the following:

Possible site: 18
>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -7.06 Transmembrane 5 - 21 (3 - 23)
INTEGRAL Likelihood = -2.28 Transmembrane 191 - 207 (189 - 207)

----- Final Results -----

bacterial membrane --- Certainty=0.3624(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8755> and protein <SEQ ID 8756> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Cread: 2
McG: Discrim Score: 8.68
Ovi: signal Score (-7.5): -5.71

-1377-

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 2 value: -7.06 threshold: 0.0
 INTEGRAL Likelihood = -7.06 Transmembrane 5 - 21 (3 - 23)
 INTEGRAL Likelihood = -2.28 Transmembrane 191 - 207 (189 - 207)
 PERIPHERAL Likelihood = 4.35 142
 modified ALOM score: 1.91

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.3824 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8756 (GBS259) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 4; MW 54kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1229

A DNA sequence (GBSx1305) was identified in *S.agalactiae* <SEQ ID 3815> which encodes the amino acid sequence <SEQ ID 3816>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 136 - 152 (135 - 152)

----- Final Results -----
 bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD47593 GB:AF140704 Vexp2 [Streptococcus pneumoniae]
 Identities = 152/212 (55%), Positives = 152/212 (71%)

Query: 1 MLELNIAIYRYKGMNDNKLTENINYSFQSGVFYITLQNSGGSGKTTLLSLMAGLDSPTGQV 60
 +L+L++ YRYK L INY+F+ G FY+I+G SG+GK+TLLSL+AGLDSPT EG +
 Sbjct: 3 LLIQLQDVYTRYKNTAEAVLYQINYNFEPGKIFYISIGSGAGKSTLLSLAGLDSPTVEGSI 62

Query: 61 LFNKDKIEAGYAGHRKKNIALVPQNYNLIDYLTPLENVQLVKPTADKQLLLDLGLKEDM 120
 LF +DI++ GY+ HR +I+LVFQNYNL+DYLT+PLEN++LV A K LL+LGL E
 Sbjct: 63 LFQGEIDIRKGYSYHRMHLISLVFQNYNLIDYLSPLENIRLVNKKASKNTLLSLGLDESQ 122

Query: 121 LTRNLRLLSGGQQORVAIARALVVGTPAILLDEPTGNLDFDISRDITMLKDFAHKRC 180
 +RN+L+L+SGGQQORVAIAR+LV P IL DEPTGNLD +DI LK A K +C
 Sbjct: 123 IGRNVQLSGGQQORVAIARSLVSEAPVILADEPTGNLDPKTAGDIVEKLKSLAQKTGKC 182

Query: 183 VIVVTHSRKTAHMADTALQLIGDLNKLKSKES 212
 VI+VTHS+E+A +D I+L L E S
 Sbjct: 183 VIVVTHSKEVAQASDITTELEKDKKLTETRNIS 214

SEQ ID 3816 (GBS363) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 5; MW 28kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 10; MW 53kDa).

GBS363-GST was purified as shown in Figure 216, lane 9.

-1378-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1230

A DNA sequence (GBSx1306) was identified in *S.agalactiae* <SEQ ID 3817> which encodes the amino acid sequence <SEQ ID 3818>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood = -14.97  Transmembrane    71 - 87 ( 66 - 97)
    INTEGRAL    Likelihood = -3.61  Transmembrane    2 - 18 ( 1 - 18)

----- Final Results -----
        bacterial membrane --- Certainty=0.6986(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1231

A DNA sequence (GBSx1307) was identified in *S.agalactiae* <SEQ ID 3819> which encodes the amino acid sequence <SEQ ID 3820>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

25  Possible site: 45
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.1986(Affirmative) < succ>
30  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1232

A DNA sequence (GBSx1308) was identified in *S.agalactiae* <SEQ ID 3821> which encodes the amino acid sequence <SEQ ID 3822>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

45  Possible site: 34
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -6.05  Transmembrane    22 - 38 ( 17 - 39)

----- Final Results -----
        bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


-1379-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]
 Identities = 39/153 (25%), Positives = 67/153 (43%), Gaps = 9/153 (5%)

Query: 3 LFKRSFLVSGRIKRSITLFLVCLMLVASTLLSGIAVKNAGLTA-KKTPSRQQTGSILHISS 61
 + +F YVVKR KSI +F+ + L+AS + G+++K A A +TF T S +
 Sbjct: 1 MLNAPAYVTRKFFKSIIVIFLILIMASLSLWGLSIGATAKASQRTFKNTINS-PSMQI 59

Query: 62 DSTDLVGDGSGGRIPEKAIYNIAENPNVKNVNNIMAYAGLTSEKMTRENDKEQYKE- 120
 + + G G+G I + I I N ++ + A LT ++ P K+
 Sbjct: 60 NRRVNGQTPRGAGNKGEDIKKIKTENKATESYVVKRINAIGDLTGVDLIETPETKKNLAD 119

Query: 121 -----CNVLQVHGNSYSDDPKYTAGMISLKGK 147
 L + G + S + K + +G L G
 Sbjct: 120 RAKRPGSLMITGVNDSKEDKPFVSGSYKLVEG 152

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1233

A DNA sequence (GBSx1309) was identified in *S.agalactiae* <SEQ ID 3823> which encodes the amino acid sequence <SEQ ID 3824>. Analysis of this protein sequence reveals the following:

25 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -15.76	Transmembrane	295 - 311 (287 - 317)
INTEGRAL	Likelihood = -7.59	Transmembrane	49 - 65 (46 - 69)
INTEGRAL	Likelihood = -6.90	Transmembrane	340 - 356 (339 - 362)
INTEGRAL	Likelihood = -5.57	Transmembrane	411 - 427 (404 - 430)

30 ----- Final Results -----

bacterial membrane	---	Certainty=0.7105 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

35

A related GBS nucleic acid sequence <SEQ ID 9695> which encodes amino acid sequence <SEQ ID 9696> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB12182 GB:Z99106 similar to transporter [Bacillus subtilis]
 Identities = 95/370 (25%), Positives = 167/370 (44%), Gaps = 41/370 (11%)

Query: 109 ESYEASLSDVSGSRKLSVSPYNS-----KEENQVTLAQGTEDLRAPQTKALNLK 160
 +++E+S S D S S + NS + + + + G ST + F +
 Sbjct: 115 DATESSSSSSSSSSSSSSSNAKNSQGGGQGGPQVQADLSIEGVISTALVDDPSDGSKIT 174

Query: 161 KGSFLAADNT--KQVLVPLKLAQKNHLSVGNKRLGK---ENVT----IAGIYDANSA-- 209
 G + + + K ++ LA++N LGGV+ + E+ T I GIY S+
 Sbjct: 175 DGRAITKSDVGKVTVINETLAEENDLGVGDSITTESATDEDTVKIKLVGIYKTTSSGD 234

50 Query: 210 -KSKNTFNPNIDNTLIAQATLVKRIKQKGYQTV---AVRLSKRLVDTVIGNIKKNPLD 265
 ++N N N L T + T + + D + +DT ++ K+ +D
 Sbjct: 235 DQAQNFSPFLNPKLATPYTATAALGSDDYKNTIDSAVYMDAIGMOTFVFAAKKTSID 294

55 Query: 266 PGKLDVQTAKEPFYDGSYRNITELHRLVGRILLVSLVAMAILVVM/LFWINNRKIKTGIL 325
 F + T + Y IE + ++ +VH+ IL ++ I R E G+L
 Sbjct: 295 FDTYTLATNDOLYQGMVGTENVASPSKRVVYLVSAGVAGILGLVW4SIREKIKTEGVL 354

```

Query: 326 LAIGTAKTTHIGHYLVFLVLAAGAAFTLISIGGVGKTFAGGLLSQV----- 373
      +ATG+ +++++I+G L E+L+VA A L+ + G + L+SQ
Sbjct: 355 MAIGKEKWLIGQFLTEILLVAVLAIGLAVTGNVANLQGLQSQSSSTDSQTAS 414

5 Query: 374 -----GSSSVQIVCNSSILIDRINDLAVSVGVFRLVAQGLILCFVAVLSSYSI 427
      GG+ + + +SS +D ID+H V+V +D+ L G L+ +A +D L S+L
Sbjct: 415 GQMPGGGGGSGKMPGSSSNVDVLSLNAVSMNDMLLGGITGLTALATLPSISVL 474

10 Query: 428 LKIQPKQLR 437
      +L PK L+L+
Sbjct: 475 RLKPKTLTK 484

```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8757> and protein <SEQ ID 8758> were also identified. Analysis of this
15 protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 9
McG: Discrim Score: 1.50
GVH: Signal Score (-7.5): -8.43
Possible site: 39
20 >>> Seems to have an uncleavable N-term signal seq
ALOM program count: 4 value: -15.76 threshold: 0.0
    INTEGRAL Likelihood = -15.76 Transmembrane 295 - 311 ( 287 - 317)
    INTEGRAL Likelihood = -7.59 Transmembrane 49 - 65 ( 46 - 69)
    INTEGRAL Likelihood = -6.90 Transmembrane 340 - 356 ( 339 - 362)
25 INTEGRAL Likelihood = -5.57 Transmembrane 411 - 427 ( 404 - 430)
    PERIPHERAL Likelihood = 3.45 386
modified ALOM score: 3.65

*** Reasoning Step: 3

----- Final Results -----
    bacterial membrane --- Certainty=0.7305 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF0687(421 - 1611 of 1917)
RGAD|108957|BS0375(11 - 484 of 486) hypothetical protein [Bacillus subtilis]
CMN1|NT01BS0429 membrane transport protein GP|1805444[dbj|BAA09006.1]|D50453 homologue of
hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus
[Bacillus subtilis] GP|2632675[emb|CAB12182.1]|299106 similar to transporter [Bacillus
subtilis] P|R|F69762|F69762 transporter homolog yclI - Bacillus subtilis
%Match = 8.6
%Identity = 28.7 %Similarity = 52.2
Matches = 117 Mismatches = 184 Conservative Sub.s = 96

      312       342       372       402       432       462       492       522
VL*NH*LIIIVDEVREYITTSIVLEILIKRKQKGVINWTLGIAYLRKKQMKVTFLFLVLTITGTCLISMSIQHSLEK
                                : | : | : || : : | : : | : : |
MNFIKRAFWNMKAAGKKTLLQLFPVTIVCVFLSGLIAQAQAK
                               10        20        30        40

      543       573       603                                     624       654
N---ILTKGKSLYLTSKEKAYWPEQAYEALK-----ARMVSVEASESID
      : : : | : : : | : : : | : : : | : : : | : : : |
SSELARELGGSVTLGVDRQKNQGGQSEKRTFBSTPIKVSDANIKLAALDHVKSYNYTTSASANGNFDAESSSSD
                        60         70         80         90        100        110        120

      684       720       750       780       807       834       864
VGSRLSKVSVPNS-----KEENVTLAQYSTEDLRFAQTALVIKSGSHLA-AMTKQV-LVFELKALQKHLSVG
      : : : | : : : | : : : | : : : | : : : | : : : | : : : |
SSSSSSSNANSNGCGGQGQPQNVQDLGISVSLTAFLVDPSDDSKITDGRATFSDDGVKKTVVINETLAEENDLSVG
                    140          150          160          170          180          190          200

```

-1381-

```

      885      903      954      978      1008      1065
NKLRL---GKENVTI---AGIYDANSA---KSKNTFNNIDNTLIA--QNTLVKISKQKGYQTVAVR--LSDKRLDVTV
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 DBITTESATDEDTTVKLGIVGIYKTTSSGDDQANFSFLNPNYKLYTPPTATAALKGDDYKNTIDSAVYYMDAKNMDTF
      220      230      240      250      260      270      280

1095      1125      1155      1185      1215      1245      1275      1305
IQNIKQWPLDFGKLDVQTAKEFYGDSYRNIEITLHRLVGRILLVSLVAMAILAVMLTFWINNRIRKRTGLLAIGKTKFEI
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 VKAAKTSIDFDTYTLANTDQLYQOMVPIENVASFKNVVLVSVAAGVILGLIVMSIRERKYENGVLMAIGEKRNKL
      300      310      320      330      340      350      360

1335      1365      1395      1431      1461      1491
IGHYLIEVLVAGAAFTLSITGGVFLGKTFAGLLSQV-----NGGVSSQIVQNSSLIADIRIDLAV
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
15 IQQLTEILLVAVIAIGLASVIGNLVANQLGNOLLSSQISSSTDTSTASQMPGGGGGAGGKMGFSSSSNNVDIDSLV
      380      390      400      410      420      430      440

1521      1551      1581      1611      1641      1671      1701      1731
SVGMVDVRLVYAGALICLAVVLSSYSILKQPKLQSRMS*EVNNMLFKRSFLVSRKKRSITLFLVCLINASTLIS
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
20 AVSNMELMLGGIGILIAITLTLSPISVLRLHPKTLITKQE
      460      470      480

```

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1234

A DNA sequence (GBSx1310) was identified in *S.agalactiae* <SEQ ID 3825> which encodes the amino acid sequence <SEQ ID 3826>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
35      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40 >GP:CAB11993 GB:Z39105 ybdG [Bacillus subtilis]
   Identities = 66/224 (29%), Positives = 102/224 (45%), Gaps = 22/224 (9%)

Query: 84 IKEYGQKQVVGKGMVYTVGEGKVPVIFVPGQVTVTAHQYHNLSNLSKTHKVVVPEP 143
      +K G V+V GKQENVY G GK VP+ G G ++ L S SK +K+ VV+
Sbjct: 41 LKSGKTVVDVGGKGMVYVGGSGKDTFFVMSGG3IAAPAYEMKGLYSKFSKENKIAVDR 100

45 Query: 144 FGSGLSDVDQPRNLANITSDIHEALQKVGITGKVIASHSIGGVYALKYISTYKPEVLG 203
      G G S+V R++ + +AL K G Y++ HSI G+ A+ + YPKK+
Sbjct: 101 AGYGYSEVSDDDRDIIDTLEQTRKALMKSCKNPFIYLMPHSISGTEAMVWQYKPKKA 160

50 Query: 204 LIGLDTSTP-----GMEGGKQVDF-----AAPVLKELPKPKVSDDDN 241
      +I +D P G++ K F +A E+ + +D+
Sbjct: 161 LIANDIGLQQQVYTKLGSVDRLKVRGPHLLTSGFHRFIPSAVYNPEVIRQSFILTDEEK 220

Query: 242 AQFFAIGHKLIANNMKEEAKNSNMINESANYKIPKGIAMYL 285
      + AI K N+M+ E S ++S N PK P+ L
55 Sbjct: 221 EIKYKINFQFFNADMHEHLLQSTONGSKSVNLPAPKETFVLIL 264

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1382-

SEQ ID 3826 (GBS121) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 9; MW 40kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 6; MW 65kDa).

GBS121-GST was purified as shown in Figure 198, lane 6.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1235

A DNA sequence (GBSx1311) was identified in *S.agalactiae* <SEQ ID 3827> which encodes the amino acid sequence <SEQ ID 3828>. Analysis of this protein sequence reveals the following:

10 Possible site: 33
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 15 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8759> which encodes amino acid sequence <SEQ ID 8760> was also identified. Analysis of this protein sequence reveals the following:

20 Lipop: Possible site: -1 Crend: 8
 MoG: Discrim Score: 3.70
 GVH: Signal Score (-7.5): -0.0600004
 Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.
 25 ALOM program count: 0 value: 8.01 threshold: 0.0
 PERIPHERAL Likelihood = 8.01 167
 modified ALOM score: -2.10
 *** Reasoning Step: 3
 30 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8760 (GBS60) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 7; MW 38.6kDa).

GBS60-His was purified as shown in Figure 193, lane 3.

- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1236

A DNA sequence (GBSx1312) was identified in *S.agalactiae* <SEQ ID 3829> which encodes the amino acid sequence <SEQ ID 3830>. This protein is predicted to be unnamed protein product. Analysis of this
 45 protein sequence reveals the following:

Possible site: 21
 >>> May be a lipoprotein

-1384-

GBS21L was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 8-10; MW 66.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 11; MW 41.5kDa) and in Figure 180 (lane 6; MW 41kDa). GBS21L-His was purified as shown in Figure 232 (lanes 3 & 4)

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1237

- A DNA sequence (GBSx1313) was identified in *S.agalactiae* <SEQ ID 3831> which encodes the amino acid sequence <SEQ ID 3832>. This protein is predicted to be endopeptidase O. Analysis of this protein
10 sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15      bacterial cytoplasm --- Certainty=0.3854(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

- ```
>GP:AAF67832 GB:AF179267 endopeptidase PepO2 [Lactococcus lactis]
Identities = 21/36 (58%), Positives = 26/36 (71%)

Query: 1 MRANIFVRNPFQBFYDAFGVKKGDSMYLKPEKRLTLW 36
 +RANIF N +EFY+ F VK+ D MY PEKRI +W
25 Sbjct: 592 LRANIPPTNLEEFYETFDVKETDQVYRAPEKRLKIW 627
```

There is also some homology to SEQ ID 2384:

- ```
Identities = 13/36 (36%), Positives = 25/36 (69%)

30  Query: 1  MRANIFVRNPFQBFYDAFGVKKGDSMYLKPEKRLTLW 36
      +R N+ + NF F++ F +K+GD+M+ P+ R+ +W
      Sbjct: 596 LRNVVLITNFDAPHETFDIKEGDMMWRAPKDRVIW 631
```

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1238

A DNA sequence (GBSx1314) was identified in *S.agalactiae* <SEQ ID 3833> which encodes the amino acid sequence <SEQ ID 3834>. This protein is predicted to be endopeptidase O. Analysis of this protein
sequence reveals the following:

- ```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.3801(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

- ```
>GP:AAAL6168 GB:LI8760 endopeptidase [Lactococcus lactis]
Identities = 118/268 (44%), Positives = 174/268 (64%), Gaps = 6/268 (2%)
50
```

-1385-

Query: 1 MGDYGGKYFGAEAKKDVEHMAKKIINVYKTRLKNNTWLSENKAKAKKLDNMRIMIGY 60
 +G +YGGKYFGAEAK DV+ M ++I VY+ RL N WLS+ T AI+KLD + IG+
 5 Sbjct: 321 IGLFYGGKYFGAEAKADVKRMVTIMIKVYQVRLSKNWLSCQETAEKALTEKLDATTPFIGF 380

Query: 61 PEDYFDLYRQYQFDSKASPFENNNDYRKLNNKKTPEEPNQNQRHHWQMSANAVNAYNDP 120
 P+ P+Y + + S S +E+ + K+ +TFE+P++ + W M A+ VNAY P
 10 Sbjct: 381 POKLPEIYSRLKTTGSLYEDALKPKDKIITARTFEKFSREDVDKTSWHPAHMNVNAYTSP 439

Query: 121 NTSIVFPAAALFQSPLYDKTKTVSONYGAIGAIIGHSISHSPDINGMKYDEKGNLHDWMT 180
 ++N+IVFPAAAL Q+P Y ++ SONYG IGA+I HEISH+FD NG ++D++GNL+ WW
 10 Sbjct: 440 DSNITIVFPAAALQAPPYSLQSSSNYGGIGAVIAHEISHAFDNNQAQFDKRCNLNKWML 499

Query: 181 KEDLKHYYKKTKQAMIDQMDGLKADGGKVDGKLTAEINIADNGGVNASLKALEKTIKTIK 240
 ED ++K+ + MI +DGL+ + G +GKL ++ENIAD GG+ A+L A K EK +K
 15 Sbjct: 500 DEIDYAFERKQKEMIALFDGVETRAGFANGKLTIVSENIADQGGITAAI/TAADKEKQVDLK 559

Query: 241 NFLNHGQVPGVKQPKNKVSPQFSQMF 268
 F + K + K S +F QM +
 20 Sbjct: 560 AFFSQN-----AKIWRMKASKEFQQML 582

There is also homology to SEQ ID 2384:

Identities = 110/253 (43%), Positives = 161/253 (63%), Gaps = 1/253 (0%)

Query: 1 MGDYGGKYFGAEAKKDVEHMAKKIINVYKTRLKNNTWLSENKAKAKKLDNMRIMIGY 60
 +G +Y + F AK DVE ++I VYK+RL+ WL+ T+ AI KL+ + IGY
 25 Sbjct: 324 IGLNYAGQKFSPEAKADVESKVMIEVYKSRLETADWLAPATREKAITKLVNITPHIGY 383

Query: 61 PEDYFDLYRQYQFDSKASPFENNNDYRKLNNKKTPEEPNQNQRHHWQMSANAVNAYNDP 120
 PE P+ Y + D S EN N K++ T+ ++N+ R W M A+ VNAY D
 30 Sbjct: 384 PEKLPETYAKKVIDESLSLVENQNLAKITIAHTMSKCNKPFVDSRSEWHPAHLDVNYTDL 443

Query: 121 NTSIVFPAAALFQSPLYDKTKTVSONYGAIGAIIGHSISHSPDINGMKYDEKGNLHDWMT 180
 N IVFPAAAL Q P Y ++ S NYG IGA+I HEISH+FD NG +DE G+L+DWM
 35 Sbjct: 444 CQNQIVFPAAALQAPPYSLDQSSSNYGGIGAVIAHEISHAFDNTNGASFDHSGSLNDWMT 503

Query: 181 KEDLKHYYKKTKQAMIDQMDGLKADGGKVDGKLTAEINIADNGGVNASLKALEKTIKTIK 240
 +ED +K++T ++ Q+DGL+ + G KV+GKLT++EN+AD GGV +LEA ++E+ ++
 40 Sbjct: 504 QEDYAAPKERTDKIVAQFDGLSHGAKVNGKLTIVSENVADLGGVCALEAAQSEDFDSAR 563

Query: 241 N-FLNHGQVPGVK 252
 + F+N ++ +K
 45 Sbjct: 564 DFFINFATIRMK 576

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1239

A DNA sequence (GBSx1315) was identified in *S. agalactiae* <SEQ ID 3835> which encodes the amino acid sequence <SEQ ID 3836>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 55 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9691> which encodes amino acid sequence <SEQ ID 9692> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1386-

>GP: AAC35997 GB: AF019410 endopeptidase O [*Lactobacillus helveticus*]
 Identities = 85/315 (26%), Positives = 146/315 (45%), Gaps = 8/315 (2%)

Query: 46 NVSPRENTLYRAVDNWLANTKLGQGSQTSVNSFSEIEDKLGQLLVSDMAKMASGRIKTTN- 104
 N P++NLV AVN WL+ ++ QTS +B++ K+++ ++ D A +ASGK + +
 Sbjct: 20 NAKPQNTLYLAVNSRWLSKARIPADQTSAGVNTSEIDIKIEKRMMDQFADIASGKEKMPDI 79

Query: 105 DRQKKMVAAYYKQGMDFKTRDNGKLKPLKPVLQKLEAVSSMKDPQSLAHDPMNSGFVLPFG 164
 + K +A YK +F RD P++ LQK+ + + F+ A + N + LDP
 Sbjct: 80 RDPFKALALYKIAFNFDKDAKANPIQNDLQKILLDINFDKFKDINATELPMGPTALPPV 139

Query: 165 LTVRTNARDNSQKQLVLRQAPALLRSFDQYKKGNGEBAKLISAYRTSMALLKQAGKSNIT 224
 V+ + + L L YK E + L + IL+ AG
 Sbjct: 140 FDDVADMDNTDENVLHFGGPSTFIPDTTYK--TPEAKLLDILEKQSLINLEWAGIGKE 197

Query: 225 EDRKLVRQAIADFELLSEKTIQVDQSKITASSETAAGRYNPESMETVHNYAKEFDKELIE 284
 E R V+ A+AFD+ LS+ K T E A YNP S+ K PD + ++
 Sbjct: 198 EARVTVQNALAFDQKLSKV-----KSTEESSDYAATYINPVSLTSLFAKFKSFMAFLK 252

Query: 285 KLVGPTNKAVNVEDKTYFKQVNDVINSKQLANMKAMNMISMLVDQSDFLGEQNQAASAF 344
 ++ + V V + + +++IN +K NM++ + +L + R AA F
 Sbjct: 253 TILPEKVERVIVMEPRFLHMADELINPANDFEIKGNWLVKYINSVAKYLSQDFRAAAFFP 312

Query: 345 KQVASGLTQIESKEK 359
 SG ++ S+ K
 Sbjct: 313 NQAIISGTPELPSQIK 327

A related GBS gene <SEQ ID 8763> and protein <SEQ ID 8764> were also identified. Analysis of this protein sequence reveals the following:

30 Lipop: Possible site: -1 Crend: 10
 MoG: Discrim Score: 5.41
 GvH: Signal Score (-7.5): -1.39
 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 35 ALON program count: 0 value: 2.76 threshold: 0.0
 PERIPHERAL Likelihood = 2.76 151
 modified ALON score: -1.05

40 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8764 (GBS12) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 7; MW 65kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 3; MW 39kDa).

The GST-fusion protein was purified as shown in Figure 189, lane 4.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1240

A DNA sequence (GBSx1317) was identified in *S. agalactiae* <SEQ ID 3839> which encodes the amino acid sequence <SEQ ID 3840>. Analysis of this protein sequence reveals the following:

55 Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 301 - 317 (299 - 317)

-1387-

----- Final Results -----

bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA842180 GB:A67181 unnamed protein product [unidentified]
 Identitles = 245/771 (31%), Positives = 410/771 (52%), Gaps = 80/771 (10%)

10 Query: 22 VRVVEFNKESILDYATEQKKVIAQLAQADVEKKLQSIKOBQDKVLAKNIEKSVHFDSSKV 81
 VRVIV NK + D+ ++ + A + + + +E+ +K Q+K+K +E+ +KV
 Sbjct: 97 VRVIVSLNKSAAFDHTSKPTGSAASVK-- --IQASDQVKDQGEKVITQVEE---ITGNKV 151

15 Query: 82 KR-YDAINGVALDIQAQIEKLTADVRVIVSQEYVQTKPLSSSQGLIGLPEVWNN 140
 +R + ++N ++D+ +I+K+K + V+ V + Y P S+ Q+ + +VW
 Sbjct: 152 RRQFGYLVNFAISIDMDLDDIDKVKDLQVKVIVPVKVY-- --HPTDSADQMAQVQVQWQE 206

20 Query: 141 SQYKSGEVTVAVIDSGVDFKHQALKIKEPRAKYNKTSIE---KLIHEKNGKGFYSEK 196
 + KGE V+++ID+G+D HQ LK+ +K+ +E KLH GK+Y+EK
 Sbjct: 209 QRLKGGHVISIIDTIDSHQQLKDSGVSSTALSKSEVSDSKLGH---GKYITK 263

25 Query: 197 VPKGYITVDYDNENLKDS-YGVMEKMHNTGIVGANDDNKLYGVAPNAQILANKVFSDEQQ 255
 VPKGYI D ND + D+ G MEG HV GI GAN ++ GVAP+AQ+LANKVFS++ +
 Sbjct: 264 VPKGYIADKNDQIVDNGCOEHHGQHVAGTAGNG---QVKGVAQQLLANKVFSNNK 320

30 Query: 256 NPPTFTDVLKALDDAILKADVPVNSLGTGAGFVHEKDYPELEVIAPAKAGIIVIAVA 315
 N + D + A++D+ L ADV+NSLGS+ V G P+ +A+A+AG++ ++
 Sbjct: 321 NSGAYDDDIISAIEDSVKLGDVNNLSGSVSSDV--GPSDPCQAVAKASEAGVINVIS 376

35 Query: 376 AGNE---GNITDGTNYGVKPLAENYDIALANPALDNTLAVASMENLKHIAHVLKFK-- 370
 AGN G+ DGN +E ++ P + + L VAS EN K +K +
 Sbjct: 379 AGNSGVAGSTADGNFVNTGTSE--LSTVGTGVTFDALTVASAENKVTVDIVDELG 435

40 Query: 371 -----DKSGTGVTEVINLHVAPNASKTIIGLAVDLGAGAPSELS--KHPDLGKIA 420
 + K +VT + + + K + VD+G G + + K + +G+ +A
 Sbjct: 436 GVTSSNSSELGAAQVTTQLESNYSVLTKLKL---VDMGLGADGYTAERKAEVKGOLA 492

45 Query: 421 MLEIFPEINKSGFLEKQVATIKINPAILLNNNAKVKDILGSQLLVESEAKFNIAITR 480
 +++ + F KV A I++YN+ D L S L + ++
 Sbjct: 493 VVK----RGAYTFSAKVANAKAAGAAGIIVTYNSE--DGLLSMSLDKTPPTLGMKADG 546

50 Query: 481 STY---NNIKNNNNKIITILTERQADINSLAQSSYSWSGPTPDLRLKPRITAPGGHI 536
 + ++ + K T L IDNS AG+ +S +SWGPT+L KPRITAPGGI
 Sbjct: 547 KFWLKQKKVRSRLKFTAL-----IDNSRAGKMSDFTSWGPTPELDFKPRITAPGGKI 601

55 Query: 537 FSTVEDNOYADKSGTSMAPQVAGAAVLKQYITDKIPV--DNAADFILKLLNTAQPI 594
 +S DN+Y SGTSMAP VAG+ A++ Q I + + + + F K MVT+ P+
 Sbjct: 602 YSLANDNKYQMSGTSMASFFVAGSEALLQGIKKQGLNLSGSELVQFAKNSAVNTSHV 661

60 Query: 595 IN-KQSKDGKTPYFVFRQGSQGMNLAKALVTVVATVIGTNDNNAAGKLEEL-KEKFK 652
 + + +K+ +P R+QSG +N+ A+ TV N+G L+E+ ++ P
 Sbjct: 662 YDTBHTKEIISP---RQSGSEINVKDAINNTVEKKA---NGNAAALKEIGRQTF 713

Query: 653 KARILLNPFNGKINTKYIISSEA---IADPVDEKGFQKNSBHLVSKKADAVTKVTVKAGK 710
 K + L N GK +TY + + + K +++ +V + T KVT+ G+
 Sbjct: 714 K--VILINHGKKAQYAVDNYGGPYTQATKSGEYIDTK-IVKQQLATTEPKVTQVQE 770

Query: 711 TLAVDLDDVDSAEALTRNNFLLEGYINLK-DTEGVADLHLFLPLFGSGWTE 760
 +VD+ + + R NF+BGY+ + + +L LP++GP+GS+++
 Sbjct: 771 --SVDVSFILTLPYSPQNFVBSYVGFRAKDQATPNLVLFYMGFFGYSQ 819

A related GBS gene <SEQ ID 8767> and protein <SEQ ID 8768> were also identified. Analysis of this protein sequence reveals the following:

65 Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -8.37

-1388-

GvH: Signal Score (-7,5): -6.06

Possible site: 15

```
>>> Seems to have no N-terminal signal sequence
```

ALOM program count: 1 value: -1.75 threshold: 0.0

INTEGRAL	Likelihood = -1.75	Transmembrane	301 - 317 (299 - 317)
----------	--------------------	---------------	------------------------

PERIPHERAL Likelihood = 1.75 614

modified ALOM score: 0.85

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.1702(Affirmative) < succ>

```
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00677 (358 - 3159 of 3255)

EGAD|139899|149200(95 - 1541 of 1946) prtB protein {*Lactobacillus delbrueckii*}

GP 1381114 gp AAC41529.1 L48487	proteinase precursor	{Lactobacillus delbrueckii}
---------------------------------	----------------------	-----------------------------

PIR JC6032 JC6032	lactocenin (EC 3.4.21.96)	precursor	[similarity]	-	Lactobacillus
-------------------	---------------------------	-----------	--------------	---	---------------

delbrueckii subsp. bulgaricus

```
de1blueck11 8
%Match = 15.5
```

Identity = 33.3 Similarity = 54.6

Matches = 275 Mismatches = 343 Conservative Sub.s = 176

318 348 378 408 438 468 498 528
KAVTVTKPGVAZKAATPAVPKPQKVRIVFNKESILDYAYEKKTVAQLMQADVEKKLQSIOKEQDKVLKNIEKSVMH
| | | | : | : | : : : : : : : : : : | : | : | :
SKFGAAKQRQASGQAKSSKNSSVRIVSLNKAAPDHTSKPTGSANW--KKIQASDVQDKQGVKIKQVEE---I

90 100 110 120 130 140

555 585 615 645 675 705 735 765

TGNKVRRRQFGYLWNAFSIDMDLDDIDKVKLDPQVKVNTVPVKVYHPT---DESADQMAQVDVVWQEQLKGEGMVISIIDI

160 170 180 190 200 210 220

795	825	855	885	915	942	972	1002
-----	-----	-----	-----	-----	-----	-----	------

GIDSSHQDLKLDGSGVSTALSKSEVESDKS-KLGHGKYYTEKVPFYGYNYADKNDQIVDNGCGEMHGQHVAGIAGAN---GC

1032 1062 1092 1122 1152 1182 1212 1242

VKGVA³¹⁰PD³¹¹QA³¹²LL³¹³AM³¹⁴KV³¹⁵PS³¹⁶NN³¹⁷AK³¹⁸NS³¹⁹GA³²⁰YD³²¹DD³²²II³²³SA³²⁴IED³²⁵SV³²⁶KL³²⁷GA³²⁸VD³²⁹IN³³⁰MS³³¹LG³³²SV³³³SS³³⁴SD³³⁵--³³⁶GPS³³⁷PD³³⁸QQ³³⁹AV³⁴⁰AK³⁴¹SE³⁴²AG³⁴³VI³⁴⁴NT³⁴⁵

1272 1302 1326 1356 1386 1416 1656

: || | | : | : : : | | | | |
 SAGNSG--VAGSTADGNPVNNTCTSELSTVGTPGVTPDALIVASAENSK-----
 390 400 410 420

1686 1716 1746 1776 1806

-----VTTDFVKDELGGVTFSSNSSELKG-AAQVTTQLSESNYSVLTKKLLVDNGLGGADDTT-----FWLKQQ
430 440 450 460 470 480

1824 1854 1884 1914 1944 1974 2004

||| |::| ::|||::| :|||::| |::| |||::| |||::|
 KKVRASRLKFGTALDINSRAGKMSDFTSWGPTPELDFKPEITAPGGKIYSLANDNKYQMSGTSMASPFVAGSEALITLQ
 570 580 590 600 610 620 630

[illegible]

SEQ ID 8768 (GBS362N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 10; MW 63.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 182 (lane 9; MW 38kDa) and in Figure 149 (lane 11 & 12; MW 38kDa). Purified GBS362N is shown in Figure 235, lanes 3 & 4

GBS362C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 14-16; MW 91kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 18; MW 66.3kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1241

A DNA sequence (GBSx1318) was identified in *S.galactiae* <SEQ ID 3841> which encodes the amino acid sequence <SEQ ID 3842>. Analysis of this protein sequence reveals the following:

possible site: 35

```
>>> Seems to have no N-terminal signal sequence
```

INTEGRAL Likelihood = -4.04 Transmembrane 21 - 37 (17 - 38)

----- Final Results -----

bacterial membrane --- Certainty=0.2614(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) \leq succ

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < success

-1390-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BA95000 GB:AB042239 Paa [Streptococcus criceti]
Identities = 55/166 (33%), Positives = 81/166 (48%), Gaps = 24/166 (14%)

5   Query: 5   KKTDKPGFRKSKVCRSLGALLGTVAIVSLATASTRIHADENTTSPTTVTKVPQVQADT 64
      K+ + PGFRKSK+ +SLOGALLGT VVS+ A A++ TTS T+ DF
      Sbjct: 2   KRKRTPGFRKSKISKSLGALLGTATVIVSV--AGQALAEADMTTSTSA-----VDT 51

10  Query: 65   TALNTSKIHSTQATTTPVEAKNKVVKSETVQSERV--MFRD-KVVERPETVKASVNS- 120
      TA+ ++T + +A + ++ Q+E + MP D E E VK++ +
      Sbjct: 52   TAVVGTETGNPATNLPEKQADSSQARASQCAEQKTSMPVDVATTELDEAVKSAEAG 111

      Query: 121 -DVSGPIITTTPTTI-----NEKTVEIPNIAQDTKKVAPKVTVIFE 159
              VSQ T T+ +EK+ EI D K A + +T E
15  Sbjct: 112   VTVSQDETVDKGTGVTGTSQEADEKSGEI---KADYSEQASTIKITIE 154
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3842 (GBS222) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 6; MW 22kDa).

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1242

A DNA sequence (GBSx1319) was identified in *S.agalactiae* <SEQ ID 3843> which encodes the amino acid sequence <SEQ ID 3844>. This protein is predicted to be CylK. Analysis of this protein sequence

- 25 reveals the following:
- ```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
30 bacterial cytoplasm --- Certainty=0.3738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1243

A DNA sequence (GBSx1320) was identified in *S.agalactiae* <SEQ ID 3845> which encodes the amino acid sequence <SEQ ID 3846>. This protein is predicted to be CylJ. Analysis of this protein sequence

- 40 reveals the following:
- ```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45   bacterial cytoplasm --- Certainty=0.1143(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9689> which encodes amino acid sequence <SEQ ID 9690> was also identified.

50

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1244

- 5 A DNA sequence (GBSx1321) was identified in *S.agalactiae* <SEQ ID 3847> which encodes the amino acid sequence <SEQ ID 3848>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0913 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1245

- 20 A DNA sequence (GBSx1322) was identified in *S.agalactiae* <SEQ ID 3849> which encodes the amino acid sequence <SEQ ID 3850>. This protein is predicted to be CylI (fabF). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
25 INTEGRAL Likelihood = -2.39 Transmembrane 721 - 737 ( 721 - 738)
   INTEGRAL Likelihood = -1.97 Transmembrane 326 - 342 ( 326 - 343)
   INTEGRAL Likelihood = -0.43 Transmembrane 534 - 550 ( 534 - 550)
```

- 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1956 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 9687> which encodes amino acid sequence <SEQ ID 9688> was also identified.

There is also homology to SEQ ID 3852.

A related GBS gene <SEQ ID 8769> and protein <SEQ ID 8770> were also identified. Analysis of this protein sequence reveals the following:

- ```
40 Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 1.08
 GvH: Signal Score (-7.5): -5.97
 Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 3 value: -2.39 threshold: 0.0
45 INTEGRAL Likelihood = -2.39 Transmembrane 712 - 728 (712 - 729)
 INTEGRAL Likelihood = -1.97 Transmembrane 317 - 333 (317 - 334)
 PERIPHERAL Likelihood = 3.45 492
 modified ALOM score: 0.98
```

\*\*\* Reasoning Step: 3

----- Final Results -----

- 5                   bacterial membrane --- Certainty=0.1956 (Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8770 (GBS361) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 4; MW 84kDa).

- 10   GBS361-His was purified as shown in Figure 213, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1246

- 15   A DNA sequence (GBSx1323) was identified in *S.agalactiae* <SEQ ID 3853> which encodes the amino acid sequence <SEQ ID 3854>. This protein is predicted to be CylF. Analysis of this protein sequence reveals the following:

Possible site: 44  
 >>> Seems to have no N-terminal signal sequence

- 20   ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3766 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25   No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1247

- 30   A DNA sequence (GBSx1324) was identified in *S.agalactiae* <SEQ ID 3855> which encodes the amino acid sequence <SEQ ID 3856>. This protein is predicted to be CylE. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

- 35   ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3498 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 40   No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1248

- 45   A DNA sequence (GBSx1325) was identified in *S.agalactiae* <SEQ ID 3857> which encodes the amino acid sequence <SEQ ID 3858>. This protein is predicted to be ABC transporter homolog CylB. Analysis of this protein sequence reveals the following:

-1393-

Possible site: 56

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -13.90 Transmembrane 271 - 287 ( 263 - 291)  
 INTEGRAL Likelihood = -10.30 Transmembrane 17 - 33 ( 14 - 43)  
 INTEGRAL Likelihood = -8.60 Transmembrane 114 - 130 ( 106 - 138)  
 INTEGRAL Likelihood = -6.69 Transmembrane 152 - 168 ( 149 - 178)  
 INTEGRAL Likelihood = -1.97 Transmembrane 186 - 202 ( 185 - 202)

----- Final Results -----

bacterial membrane --- Certainty=0.6562(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9685> which encodes amino acid sequence <SEQ ID 9686> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1249**

A DNA sequence (GBSx1326) was identified in *S.agalactiae* <SEQ ID 3859> which encodes the amino acid sequence <SEQ ID 3860>. This protein is predicted to be ABC transporter homolog CylA. Analysis of this protein sequence reveals the following:

Possible site: 57

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4122(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9683> which encodes amino acid sequence <SEQ ID 9684> was also identified. A further related GBS gene <SEQ ID 8771> and protein <SEQ ID 8772> were also identified. Analysis of this protein sequence reveals homology to membrane protein ABC transporters.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9085> which encodes the amino acid sequence <SEQ ID 9086>. An alignment of the GAS and GBS sequences follows:

Score = 85.4 bits (208), Expect = 1e-18

Identities = 68/271 (25%), Positives = 129/271 (47%), Gaps = 17/271 (6%)

Query: 39 KGFTCEHVLKIDINFDVYKGFDFPGIVGRNKGSGKSTLLKIIISQIYVERKQVTV--VDGKAVS 96  
 K + L+DIN +G F+G++G NG+GK+TL ++ Q + G + VDGK +S  
 40 Sbjct: 10 KKGYSFALRDINLIFEFGKGYFLGLGNKAGKTLTFNLLIQNFKQTSQDIKWEVDGKPLS 69  
 Query: 97 ----FIELGVGF----NPELTGRENVTNGAMLGFTKDEVDMDINDVDFAEHLHPIHQ 147  
 + +G+ F + L+T EN+ GA+ G +K +V + D+ + ++ Q  
 45 Sbjct: 70 IKDFYRHIGIVFQSNRLDDNLIVERNLISRGALYGLSKSQVRNRLKDLQTYLDTAIEKQ 129  
 Query: 148 KLNYSYSGNQVRLAFSAVAKQAGDVLILDEVLAVGDEAFORKNDYPMR-RKDSGKTTIL 206  
 K + S G G + ++ + A+ Q +L+LDE D +R D + + S T +L  
 50 Sbjct: 130 KYGSLSGGQKRKVDIARALLPOPSLLLDIPTETGLDPOSRRLDWAIAQLNQCSQMTVL 189  
 Query: 207 VTDMGAVICKYCNRAVLIEDGLVKAYGEPFDVANQYSDNTETA-EDMAAEKISVSVDIA 265  
 +TH + + C+ ++ +G + G+ Q+S N + + ++S++D  
 Sbjct: 190 ITHYLEMSA-CDVLNVLIGNIIYSGDIKSFITQHSHTINLVLIKPKKSLDLQSLADVF 243  
 55 Query: 266 KDLKVSLSINPRITTNITITFEVSYEVLKDD 296

-1394-

K ++S I D I+ E +V+ D+  
 Sbjct: 249 N--KQVLSERRIVFKD-ISVEEMMQVSDN 276

There is also homology to SEQ IDs 358, 482, 644, 686, 1832, 2529, 2720, 3882, 4028, 4104, 4280, 5090, 5498, 6034, 6500.

SEQ ID 8772 (GBS83) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 2; MW 37.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 5; MW 62.6kDa) and in Figure 28 (lane 3; MW 62.6kDa).

GBS83-GST was purified as shown in Figure 195, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1250

A DNA sequence (GBSx1327) was identified in *S.agalactiae* <SEQ ID 3861> which encodes the amino acid sequence <SEQ ID 3862>. This protein is predicted to be acyl carrier protein homolog AcpC. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3451(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1251

A DNA sequence (GBSx1328) was identified in *S.agalactiae* <SEQ ID 3863> which encodes the amino acid sequence <SEQ ID 3864>. This protein is predicted to be CylG (fabG). Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 3866.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-1395-

**Example 1252**

A DNA sequence (GBSx1329) was identified in *S.agalactiae* <SEQ ID 3867> which encodes the amino acid sequence <SEQ ID 3868>. This protein is predicted to be CylD. Analysis of this protein sequence reveals the following:

```

5 Possible site: 60
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2030(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1253**

A DNA sequence (GBSx1330) was identified in *S.agalactiae* <SEQ ID 3869> which encodes the amino acid sequence <SEQ ID 3870>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3219(Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1254**

A DNA sequence (GBSx1331) was identified in *S.agalactiae* <SEQ ID 3871> which encodes the amino acid sequence <SEQ ID 3872>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 INTEGRAL Likelihood = -8.97 Transmembrane 231 - 247 (226 - 251)
 INTEGRAL Likelihood = -7.06 Transmembrane 141 - 157 (134 - 164)
 INTEGRAL Likelihood = -2.76 Transmembrane 28 - 44 (26 - 44)
40 INTEGRAL Likelihood = -1.38 Transmembrane 123 - 139 (121 - 139)
 INTEGRAL Likelihood = -0.32 Transmembrane 199 - 215 (199 - 215)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB88836 GB:AL353832 putative integral membrane transport
 protein. [Streptomyces coelicolor A3(2)]
50 Identities = 68/264 (25%), Positives = 123/264 (45%), Gaps = 10/264 (3%)

```

-1396-

Query: 6 RNFIFIKQYMKIMEYKIDFFVGVGLVFLTQGLNLLFLNVLFOHIPSLEGWTFQQAIFI 65  
 R + + + M Y + F + G F L + + + F + + L G + + + A P +  
 5 Sbjct: 34 RAYGLAGMWIRSTMAYRTSFALTAFGNFAMTALDFVAILLMFSPVDALGGYSLPEVAFGL 93

Query: 66 YGFSLLPKGIDHIFPDNLMAALQORLIRKGEFKYLITRPSLPHVLVETFOVDALGELLV 125  
 YG S + G + L + + LG + R + R G D L R P L V + F + LG + +  
 Sbjct: 94 YGLSGVSLGADLAIGSMERLGR -VRDGTLDTLVRPAPVLAQVADRFALERLGRVVG 152

Query: 126 GPILL--STTVSSISWTVPKVLLFIPIIPFATLIYTSKIATSSIAFWKQSGAVIYIF- 182  
 G + + L + V I WT KVL L + I + + A + F + + V F  
 10 Sbjct: 153 GLLVGLVALVVDIDWTAARVLLPVALISGAGIFCAVFAAGAPQFAAQDASEVANAF 212

Query: 183 YMFNDFAKYFVAIYNNLLRWIISFVIFPAFTAYYPAFLQRRNVYFNIGSVI----LI 237  
 Y + YP + + L + FV + P AF + P + Y L R + + G + L  
 15 Sbjct: 213 YGGTMIQYPTTFALDLVRGATFVLPLAFVNNLPASTVL-GRPYELDLQWVAFTPEPLA 271

Query: 238 SLISFMVSLILMHWKGVYVESAGS 261  
 + + + W G + Y S GS  
 20 Sbjct: 272 AAACALAGLAWRAGLRSYRSTGS 295

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3873> which encodes the amino acid sequence <SEQ ID 3874>. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence  
 25 INTEGRAL Likelihood = -8.86 Transmembrane 227 - 243 ( 225 - 251)  
 INTEGRAL Likelihood = -7.22 Transmembrane 141 - 157 ( 133 - 164)  
 INTEGRAL Likelihood = -6.37 Transmembrane 123 - 139 ( 114 - 140)  
 INTEGRAL Likelihood = -2.97 Transmembrane 26 - 42 ( 26 - 49)

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4545 (Affirmative) < succo>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succo>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succo>

35

The protein has homology with the following sequences in the databases:

>GP: CAB88836 GB: AL353832 putative integral membrane transport  
 protein. [Streptomyces coelicolor A3(2)]  
 Identities = 69/262 (26%), Positives = 125/262 (47%), Gaps = 10/262 (3%)  
 40 Query: 8 RAIFIKQYLKQIMEYKIDFFVGVGLVFLTQGLNLLFLNVLFOHIPSLEGWTFEQIAYIG 67  
 + + + M Y + F + G F L + + + F + + L G + + + A F + Y G  
 Sbjct: 36 RAYGLAGMWIRSTMAYRTSFALTAFGNFAMTALDFVAILLMFSPVDALGGYSLPEVAFGL 95

45 Query: 68 FCLIPKIDHILFPDNLMAALQORLIRKGEFKYLITRPSLPHVLVETFOVDALGELLV 127  
 + G + L + + LG + R VR G D L R P L V + F + LG + + G +  
 Sbjct: 96 YGLSGVSLGADLAIGSMERLGR -VRDGTLDTLVRPAPVLAQVADRFALERLGRVVG 154

50 Query: 126 ILL--VITAGSIWVLKPKVLLFILVIFPAFTAYYPAFLQRRNVYFNIGSVI----VYSL 239  
 + + L I WT KVL L + + I + + A + F + + V F Y  
 Sbjct: 153 GLLVGLVALVVDIDWTAARVLLPVALISGAGIFCAVFAAGAPQFAAQDASEVANAF 212

55 Query: 185 YMFNDFAKYFVAIYNNLLRWIISFVIFPAFTAYYPAFLQRRNVYFNIGSVI----VYSL 239  
 + YP + + L + F + P AF + P + Y L G + + + G V + +  
 Sbjct: 213 YGGTMIQYPTTFALDLVRGATFVLPLAFVNNLPASTVL-GRPYELDLQWVAFTPEPLA 271

Query: 240 LVIALSLKLNKGLDAYESAGS 261  
 AL + W + GL + Y S GS  
 60 Sbjct: 274 AAACALAGLAWRAGLRSYRSTGS 295

An alignment of the GAS and GBS proteins is shown below.

Identities = 208/261 (79%), Positives = 238/261 (90%)

Query: 1 MTKYQRMFIFIKQYMKIMEYKIDFFVGVGLVFLTQGLNLLFLNVLFOHIPSLEGWTFQ 60

-1397-

- M K + MH IFIKQY+KQIMEYK+DF VGVIGVFLITQGLNLLFL+VLPQHIPSLEGWTF+  
 Sbjet: 1 MAKLRGMBALFIKQILKQIMEYKDFVGVIGVFLITQGLNLLFLSVLPQHIPSLEGWTF 60
- Query: 61 QIAFIYGFSLPKGIHLHLPFONLWALQQLIRKGSFQKYLIRPISPLPHVLVETPQVDAL 120  
 5 QIAFIYGF L+PKGIDHLPFONLWALQQLIRKGSFQKYLIRPISPLPHVLVETPQVDAL 120  
 Sbjet: 61 QIAFIYGFCLIPKGIHLHLPFONLWALQQLIRKGSFQKYLIRPISPLPHVLVETPQVDAL 120
- Query: 121 GELLVGFIILLSTVSSISWTPVKVLLFTIIPFATLIYISLKIATSSIAFWTKQSGAVTY 180  
 10 GELLVG ILL TT SI WT+PKVLLFI +IIPFATLIYISLKIAT+SI+FWTKQSGAVTY 180  
 Sbjet: 121 GELLVGFIILLSTVSSISWTPVKVLLFTIIPFATLIYISLKIATSSIAFWTKQSGAVTY 180
- Query: 181 IFYMFNDFAKYPAYINWLRWISFVLPFAFTAYTPAAYFLQDRNVYFMIGQVILISLY 240  
 IFYMFNDFAKYP+IY++ LEW+ISF+IPFAFTAYTPA+YFL +++ FMIGQ+++SL+  
 Sbjet: 181 IFYMFNDFAKYPMSIYHSFLRWLSFIIIPFAFTAYTPASYFLTGQHLIFMIGQVNVVSL 240
- Query: 241 SFMWSLILNHHKGVENVYSAGS 261  
 +SL LW G++ YESAGS  
 15 Sbjet: 241 VLALSILNKKWLLDAYESAGS 261
- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1255**

A DNA sequence (GBSx1332) was identified in *S. galactiae* <SEQ ID 3875> which encodes the amino acid sequence <SEQ ID 3876>. Analysis of this protein sequence reveals the following:

- 25 Possible site: 54  
 >>> Seems to have no N-terminal signal sequence
- |          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -15.60 | Transmembrane | 147 - 163 ( 134 - 178) |
| INTEGRAL | Likelihood = -8.55  | Transmembrane | 119 - 135 ( 114 - 141) |
| INTEGRAL | Likelihood = -7.85  | Transmembrane | 238 - 254 ( 235 - 260) |
| INTEGRAL | Likelihood = -1.70  | Transmembrane | 215 - 231 ( 212 - 231) |
| INTEGRAL | Likelihood = -1.06  | Transmembrane | 61 - 77 ( 61 - 77)     |
| INTEGRAL | Likelihood = -0.22  | Transmembrane | 27 - 43 ( 27 - 43)     |
- 30 ----- Final Results -----
- 35 bacterial membrane --- Certainty=0.7241(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 40 >GP:CB88837 GB:AL353832 putative integral membrane protein.  
 [Streptomyces coelicolor A3(2)]  
 Identities = 60/271 (22%), Positives = 118/271 (43%), Gaps = 13/271 (4%)
- Query: 6 RYKPFISTIGIQGLITYRVDIFILYRIGDVGAFVAPYLMKAVFDSSSQSLIQGDSQVQ 65  
 R Y + G + TYR + + + Y + A+D Q + G+ + +  
 45 Sbjet: 7 RLIVAVAGGFRRYATYRAATAGVFNIVPGLIIVTYTLLALNDEKPK--LGSYDQAV 64
- Query: 66 LYIMS-FVTNLIAETDSSFM--IGDEVKDSIIMRLLEPVHFAASYLFMEIGSRNLIFL 122  
 + + + + L F + + + + G + + L RP +L +G L  
 50 Sbjet: 65 TFWLQALLAALAGGQGFEDIMERIRITGDVADLYRPADLQLWMLADYQVAPQLL 124
- Query: 123 SIGV-PFLIVITGVRLFTGLDLQAIVLVVYIIISIIILAFINFPNFCPSGSAVFPKNI 181  
 GV PF+ LF L + + + +++++LA + F SRF +  
 55 Sbjet: 125 GRGVVPFVPG---SLPFPVALPREVSVWAAPLVAVVLAMVGVGFAFLYLVALSAPFLWDG 180
- Query: 182 WGSNLLKNSIVAPMSGSLIPLTFPPKIVADILGLFPSSLIYTPWIIIGKYDQSGIVQA 241  
 G + F SG L+PL FP + D++ LP+SSL+ P +++G D +  
 Sbjet: 181 TGVTCMAWLAGLPCSGMLPLINVPQVLGDDVVRALPSSLLQCPADVILGADP---LGT 237
- 60 Query: 242 LLIQIFWLVWVMSQLIKKKVQLHITTCQG 272  
 L Q W + ++Ab +L+ + +CGG  
 Sbjet: 238 YLPCASWAVALLALGRLLVQSAATERVVVQCG 268

-1398-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3877> which encodes the amino acid sequence <SEQ ID 3878>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -9.18 Transmembrane 252 - 268 (248 - 277)
INTEGRAL Likelihood = -7.22 Transmembrane 161 - 177 (151 - 187)
INTEGRAL Likelihood = -6.10 Transmembrane 133 - 149 (128 - 160)
INTEGRAL Likelihood = -2.81 Transmembrane 213 - 229 (211 - 230)

----- Final Results -----
bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF11144 GB:AE002002 conserved hypothetical protein (Deinococcus radiodurans)
Identities = 56/268 (20%), Positives = 113/268 (41%), Gaps = 21/268 (7%)

Query: 15 MNSFWKRYRPFSLSGIQELITYRVNFFLYRIGDVMGAFVAYYLMKAVFDSSQSLINGFT 74
M +FW++ R + + + YR ++ + + V +W S+ ING+T
Sbjct: 1 MTHFWKRYRVLMAVSLASTLEYFASITIMVLSGTLN-LVMMLVMNTQAKSAFGSQINGYT 59

Query: 75 LSDMTFVYIIMSFTVTLTKSDGSPMIGEEVKDGSIIIMRLLRPV-----HFAASYLFMEIG 129
Y + +++ L + + +++ G++ LL P+ FAA +
Sbjct: 60 EQAFAGYFLATWLVSQLLVVWVGWELDYKIRQGTLSPELLHPIDPLMRFAAH--LITKA 117

Query: 130 FRWIVLMSVGFFFLMVLGSIKVMAGLSILQLVASSCYLVSLLLAFL---INFYFNICFG 186
FR P ++VL + + A L+ Q + Y L LA L + F + G
Sbjct: 118 FR-----LPIMLVL--LLIPAAITGAQFTSQMWAYPAVLGLALLGLCVRFMEYTLG 167

Query: 187 SSAFVFKNLGNSNLKNAIVAPMGSGLIPLAFFPKMVSIVLSPLPSSSLVYTPVMIVIGK 246
AF ++ + A G PL+F+P + + ++ PF ++ P ++ GK
Sbjct: 168 LLAFVTESSSSSGEVLMWLFYAAFGCMFAPLSFYYPGNLQTLAANTFPFYMGLGPAALLAGK 227

Query: 247 YSLSQIMVALSLQIFLWMLVMVLSQVIV 274
S ++ + + + WL VM ++ + +W
Sbjct: 228 ASGAEALRGAGVLLGMLVAMVMLVRVWV 255

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 199/268 (74%), Positives = 236/268 (87%)

Query: 5 WRRYKPFISTGIQGLITYRVDFILYRIGDVGAFVAFYLMKAVFDSSQSLIQGFLSDM 64
W+RY+PF+S GIQ LITYRV+F LYRIGDVGAFVA+YLMKAVFDSS GSLI GP LSDM
Sbjct: 19 WKRYRPFSLSGIQELITYRVNFFLYRIGDVMGAFVAYYLMKAVFDSSQSLINGFTLSDM 78

Query: 65 ILVYIIMSFTVNLITKSDGSPMIGEEVKDGSIIIMRLRPVHFAASYLFMEIGSRWMLPISI 124
YIIMSFTV LLT+IDSGFMIG-EVKDGSIIIMRLRPVHFAASYLFMEIG RW++ S+
Sbjct: 79 TFFYIIMSFTVTLITKSDGSPMIGEEVKDGSIIIMRLRPVHFAASYLFMEIGPFWMLVMSV 138

Query: 125 GVPFLLVITGVRLFLGTDLIQAIVLVVFYIISIIAFLINFFNNICFGSPAFVFNKLMGS 184
G PFL+V++G+++ G ++Q + Y++S++LAFLINF+FNICEG SAFPVFNKLMGS
Sbjct: 139 GFFPLMVLGSIKVMAGLSILQLVASSCYLVSLLLAFLINFFNNICFGSSAFPVFNKLMGS 198

Query: 185 NLLKNSLVAPMGSGLIPLTFPPKIVADILGFLPSSLIYTPVMI+IGKYDGSQIVQALL 244
NLLKN+LVAPMGSGLIPL FFPK+V+ +L FLPPSL+YTPVMI+IGKY SGI+ AL L
Sbjct: 199 NLLKNAIVAPMGSGLIPLAFFPKMVSIVLSPLPSSSLVYTPVMIVIGKYSLSQIMVALSL 258

Query: 245 QIFWLVVMVLSQIWLKRVQLHITIQGG 272
QIFWL+VMV LSQ+LWKRVQ H+TIQGG
Sbjct: 259 QIFWLVVMVLSQIWLKRVQVHLITIQGG 286

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1256

A DNA sequence (GBSx1333) was identified in *S. agalactiae* <SEQ ID 3879> which encodes the amino acid sequence <SEQ ID 3880>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2013 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9681> which encodes amino acid sequence <SEQ ID 9682> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

<GP:AAF09790 GB:AE001882 ABC transporter, ATP-binding protein  
[Deinococcus radiodurans]  
Identities = 141/331 (42%), Positives = 201/331 (60%), Gaps = 34/331 (10%)

Query: 10 MIEVSHLQENFIKTVKAPGLKGAFCFLRPEKHTEPAVKDLTFDPKQILGPTGANGAG 69  
MIEV HL K+F + AV+D++F +P G+I+G++G NKGAG  
Sbjct: 46 MIEVRLCKSPARK-----PAVQDISPISIPAGEIVGYLGPNGAG 84

Query: 70 KSTTIKMLTGILKPTSGFCRIDGKLPQENQNVKIDIGVFGQRTQLWMDLALQRTYTVL 129  
KSTTIK+LIG+L P SG R+ G +P +R + +Y +G VFGQRT L+MDL ++E+ +L  
Sbjct: 85 KSTTIKVLTLGLLVFDSGEVRVGGVFGKQGRQHVARLGAVFGQRTTLWMDLFPRESLELL 144

Query: 130 KEIVYVDPDEKFRKRNAPFLNEVLENDPIKDPVPTLSLQORMRADIAASLLHNPKVFLDE 189  
+ +Y VP F + +A E+LEL F+ P R LSLQORMRAD+AA+LLH+P++LFLDE  
Sbjct: 145 RHVYRVPAARFAENLAGPTLELLGPPFNTPARALSLQORMRADIAAALHCPPELLFLDE 204

Query: 190 PTIGLDVSVKIMIRRAITQINQEEETILLTHDLSIDQLCHRIFMIDRQGRIFDGTVS 249  
PT+GLDV K+ IR + +N E T+LLTHDL D+E+L R+ MID G+ +FDG ++  
Sbjct: 205 PTVGLDVVAKERIREFVKVNAERGVT+LLTHDLGDVERLARRVMDITGRLLFDGPIA 264

Query: 250 QKLETFGKMKTL--SFDLRPGQEHISG-SLIGKSEINIKRNDLVLDIQVDSGRYQTADII 306  
+L+ +G + L P+ P Q + +L+G+ ++ Y S A I I  
Sbjct: 265 ELQARYGGERELWVEPEKAPQALPGLTLGQDGPVRV-----YGPSCAAAAPIA 315

Query: 307 QQTALADPVRDLAWTDADIEDIIRRPYRNEL 337  
Q T A VRDL ++ +E IRR Y L  
Sbjct: 316 QVT-ALAPVRDLAVKEPEVATIRIRYBGNL 345

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3881> which encodes the amino acid sequence <SEQ ID 3882>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3315 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 272/330 (82%), Positives = 305/330 (92%)

-1400-

Query: 8 MSMEVSHLQRFNFKTVKAPGLKGAQSFAPPEKKTPEAVKDLTFDVPKQQLGFIGANG 67  
 M MEVSHLQRFN KT+K PGLKGA +SF+ P + FEAVKDL+P+VPKQQLGFIGANG  
 Sbjct: 1 MMEVSHLQRFNFKTVKAPGLKGAQSFAPPEKKTPEAVKDLTFDVPKQQLGFIGANG 60

5 Query: 68 AGKSTTIRKLTGILKPTSGPCRIDGKLQKRNQYVKDIDGVVPGQRTQLMWDALQETTY 127  
 AGKSTTIRKLTGILKPTSG+CRIT+GK+PQ+NRQ YV+DIG VPGQRTQLMWDALQETTY  
 Sbjct: 61 AGKSTTIRKLTGILKPTSGYCRINGKIPQDNRYTYVRDIDGAVPGQRTQLMWDALQETTY 120

10 Query: 128 VLKSIYDVPDEKFRKMAFLNEVLKLNDFIKDPVRTLSLQGRMRADIASLLHNPVKVLF 187  
 VLKSIYDVP+K FRKRM FLNEVL+LN+FIKDPVRTLSLQGRMRADIASLLHNPVKVLF  
 Sbjct: 121 VLKSIYDVPKAFKRMDFLNEVLNDFIKDPVRTLSLQGRMRADIASLLHNPVKVLF 180

15 Query: 188 DEPTIGLDVSVKDNIRRAITQINQEERTTILLTHDLSDIQGLCHRIPMIDRQGEIFDGT 247  
 DEPTIGLDVSVKDNIRRAITQINQEERTTILLTHDLSDIQGLC RI MID+QGEIFDGT  
 Sbjct: 181 DEPTIGLDVSVKDNIRRAITQINQEERTTILLTHDLSDIQGLCDRIIMIDKQGEIFDGT 240

20 Query: 248 VSQLKSTFGKMKTLGFDLPQGEHISSLIGKSEINKRNDLVLDIQYDSSRYQTADIIQ 307  
 V+QLK++FGKMK+LGF+L+PQGE +S +G +I ++R++L LDIQYDSSRYQTADIIQ  
 Sbjct: 241 VTQLKSGFGKMKSLSPKLPQGEQVVSQFNLDPDITVERHSLDLDIQYDSSRYQTADIIQ 300

Query: 308 QTIADPFSVGLWNTADIEDIIRRFYRNEL 337  
 +T+ADF+VRD+KMTD DIEDI+RRFYR EL  
 Sbjct: 301 KTYADFVRDVNTDVIDIEDIVRRFYRKEL 330

25

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1257

A DNA sequence (GBSx1334) was identified in *S. agalactiae* <SEQ ID 3883> which encodes the amino acid sequence <SEQ ID 3884>. This protein is predicted to be Fmt. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -9.39 Transmembrane 21 - 37 ( 8 - 39)  
 35 INTEGRAL Likelihood = -7.75 Transmembrane 360 - 376 ( 359 - 381)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4758 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 40 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8775> which encodes amino acid sequence <SEQ ID 8776> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10  
 McG: Discrim Score: 8.85  
 GVH: Signal Score (-7.5): -3.75  
 Possible site: 25  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 2 value: -9.39 threshold: 0.0  
 50 INTEGRAL Likelihood = -9.39 Transmembrane 21 - 37 ( 8 - 39)  
 INTEGRAL Likelihood = -7.75 Transmembrane 353 - 369 (352 - 374)  
 PERIPHERAL Likelihood = 4.24 92  
 modified ALOM score: 2.38

55 \*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.4758 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 60 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>



-1402-

```

5 GEVETDIENKLPMAISLQKIMTIIQRLIDQDLSEDDRLSQFFPQV---KG--SNSITTHQLLTHTSGL----REKG
 :
 LRTKKPMKMTDFFRIGSVTKFTTATVVLQVLGHNRLKLDHEDWLPGVIOGNGYDGNKIKTQELINHTSGLAEYSRSID
 100 110 120 130 140 150 160

10 807 834 864 894 924 954 978
 VKVSPYLYLN--EREQLQFCLKHY--NFWNKKSWSYYSNINFSFLGTATVIGRTYAEHLVDVVIKNPLRLDDT--CSYQSVV
 :
 VDPDTTKKSYTABELWKMGISFFPDFAPGKGWSYSGNTGYVLLGLLIEKVIGNSYAEVENRITETPELSENTFLPGNSSVI
 180 190 200 210 220 230 240

15 993 1023 1053 1083 1113 1143 1173 1203
 --NH--DLWSPMRKNGKLNKINIFNQVSTAYGAGDFPTPLNFWVIARSFSGKYFFPTDYTKHONDAISHYGGLYMH
 :
 PGTNHARGYVOP--DGASELKDVITYYN--PSAGSSAGDMISTADDLNKFFSYLLGGKLLKEQOLKMLTTPVTGKEGIDGYG
 260 270 280 290 300 310 320

```

SEQ ID 8776 (GBS61) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 3; MW 68kDa).

20 GBS61-GST was purified as shown in Figure 195, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1258

25 A DNA sequence (GBSx1335) was identified in *S.agalactiae* <SEQ ID 3887> which encodes the amino acid sequence <SEQ ID 3888>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

```

```

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2398 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1259

40 A DNA sequence (GBSx1336) was identified in *S.agalactiae* <SEQ ID 3889> which encodes the amino acid sequence <SEQ ID 3890>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -5.57 Transmembrane 16 - 32 (13 - 33)

```

```

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.3230 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

50 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.



-1403-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1260

A DNA sequence (GBSx1337) was identified in *S.agalactiae* <SEQ ID 3891> which encodes the amino acid sequence <SEQ ID 3892>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3910 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1261

A DNA sequence (GBSx1338) was identified in *S.agalactiae* <SEQ ID 3893> which encodes the amino acid sequence <SEQ ID 3894>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25 bacterial cytoplasm --- Certainty=0.4239 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1262

A DNA sequence (GBSx1339) was identified in *S.agalactiae* <SEQ ID 3895> which encodes the amino acid sequence <SEQ ID 3896>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.4349 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

45 No corresponding DNA sequence was identified in *S.pyogenes*.

-1404-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1263

A DNA sequence (GBSx1340) was identified in *S.agalactiae* <SEQ ID 3897> which encodes the amino acid sequence <SEQ ID 3898>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4962 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1264

A DNA sequence (GBSx1341) was identified in *S.agalactiae* <SEQ ID 3899> which encodes the amino acid sequence <SEQ ID 3900>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4014 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAQ38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
Identities = 23/35 (65%), Positives = 28/35 (79%)

Query: 9 LIHWEGNSGDKLIEHQTSATGWYQVDRSPSPQPKG 43
 L +HWEGNSGDKL+E CT AT WYQ+ ++ FSQ G
Sbjct: 180 LTYWEGNSGDKLLERQTRATEWYQIEKGFPSQTNG 214
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1265

A DNA sequence (GBSx1342) was identified in *S.agalactiae* <SEQ ID 3901> which encodes the amino acid sequence <SEQ ID 3902>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2036 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1266

A DNA sequence (GBSx1343) was identified in *S.galactiae* <SEQ ID 3903> which encodes the amino acid sequence <SEQ ID 3904>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10933> which encodes amino acid sequence <SEQ ID 10934> was also identified.

SEQ ID 3904 (GBS153) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 25 (lane 3; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 4; MW 47kDa).

GBS153-GST was purified as shown in Figure 198, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1267

A DNA sequence (GBSx1344) was identified in *S.galactiae* <SEQ ID 3905> which encodes the amino acid sequence <SEQ ID 3906>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2036 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1268

A DNA sequence (GBSx1345) was identified in *S.galactiae* <SEQ ID 3907> which encodes the amino acid sequence <SEQ ID 3908>. Analysis of this protein sequence reveals the following:

```

Possible site: 19

```

-1406-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2570(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA59773 GB:X85787 tsaA [Streptococcus pneumoniae]  
Identities = 18/33 (54%), Positives = 28/33 (84%)

Query: 2 LDVQSDENPAFKIFKVAKAGLSLDVFDKLVGRF 34  
+ QSD+N F+FKV+K KG++LD FD++GRF  
15 Sbjct: 320 KYQSDKNPFVFEVSKYKGIALDPFDRIIGRF 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3909> which encodes the amino acid sequence <SEQ ID 3910>. Analysis of this protein sequence reveals the following:

Possible site: 56

20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2405(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 18/34 (52%), Positives = 25/34 (72%)

30 Query: 1 MDVQSDENPAFKIFKVAKAGLSLDVFDKLVGRF 34  
+DVQSDN-P FK+ KV KKG+ L+ D+ V F  
Sbjct: 31 LDVQSDENPGFKVVKVILKSKGIVLNALEDSEVCGF 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 35 Example 1269

A DNA sequence (GBSx1346) was identified in *S.agalactiae* <SEQ ID 3911> which encodes the amino acid sequence <SEQ ID 3912>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

Possible site: 52

40 >>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -1.17 Transmembrane 169 - 185 ( 168 - 185)

----- Final Results -----

45 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAC13546 GB:AF019629 putative fimbria-associated protein  
[Actinomyces naeslundii]  
Identities = 53/109 (48%), Positives = 75/109 (68%)

55 Query: 13 IPKINQDLPTIYAGSSEDNLRQGVHLEGLISLPICGASTHAVLSQGRMPAARLFADLDK 72  
IP I+ DLP+Y G+ +D L +G+GHLGG SLP+GG T +V+G RG+ A +F +LKK+  
Sbjct: 93 IPSISLDLPVYHGTADDTLLKGLGLGTSLSFVGGEGTSTVTHRGLEAETMPINLDKV 152

Query: 73 KKQDYFYVTLNKEITLAYQVDRINVIKPSQLDAVSIERDKDYVTLITCTP 121

-1407-

K G D V R L Y+V V-EP + +A+ +EE KD +TL+TCTP  
 Sbjct: 153 KTGDSLIVEVFGVILTVRVSTIKVVEPSTEARLVRERKDLL/LVTCTP 201

There is also homology to SEQ ID 3740 and to SEQ ID 3910.

- 5 SEQ ID 3912 (GBS194) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 2; MW 24kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1270

- 10 A DNA sequence (GBSx1347) was identified in *S.agalactiae* <SEQ ID 3913> which encodes the amino acid sequence <SEQ ID 3914>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -5.15 Transmembrane 880 - 896 ( 876 - 898)  
 15 INTEGRAL Likelihood = -4.78 Transmembrane 24 - 40 ( 23 - 42)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3060 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8777> which encodes amino acid sequence <SEQ ID 8778> was also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 8  
 SRCFLG: 0  
 MoG: Length of UR: 20  
 Peak Value of UR: 2.80  
 Net Charge of CR: 5  
 MoG: Discrim Score: 10.81  
 30 GvH: Signal Score (-7.5): -3.76  
 Possible site: 29  
 >>> Seems to have an uncleavable N-term signal seq  
 Amino Acid Composition: calculated from 1  
 ALOM program count: 2 value: -5.15 threshold: 0.0  
 35 INTEGRAL Likelihood = -5.15 Transmembrane 867 - 883 ( 863 - 885)  
 INTEGRAL Likelihood = -4.78 Transmembrane 11 - 27 ( 10 - 29)  
 PERIPHERAL Likelihood = 7.58 531  
 modified ALOM score: 1.53  
 icml HYFID: 7 CFF: 0.306  
 40 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3060 (Affirmative) < succ>  
 45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>  
 LEXTG motif: 859-863

- 50 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8778 (GBS104) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 5; MW 95kDa).

GBS104-His was purified as shown in Figure 221, lane 9-10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1271

A DNA sequence (GBSx1348) was identified in *S.agalactiae* <SEQ ID 3915> which encodes the amino acid sequence <SEQ ID 3916>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -15.28 Transmembrane 257 - 273 (252 - 280)
10 INTEGRAL Likelihood = -7.11 Transmembrane 19 - 35 (16 - 39)

----- Final Results -----
 bacterial membrane --- Certainty=0.7114(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAIC13545 GB:AF019629 putative fimbria-associated protein
[Actinomyces naeslundii]
20 Identities = 79/178 (44%), Positives = 112/178 (62%), Gaps = 7/178 (3%)

Query: 65 RIALANAYNETLSRNPLL-----IDPFTSKQKEGLREYARMLEVHEQ--IGHVAIFSGIV 117
 ++ A+ADN+ LS +L + K+ +YA +L+ + + + + IPSI +
25 Sb|ct: 39 QVEQAHATNDALSAGAVLEANNHVPTGAGSKDSSLQYANILKANNEGIMARKLKIPGISL 98

Query: 118 DIPFIYAGTSETVLQKGGSHLEGTSLFVGGLSTHSVLYMHRGLPTARLPDLANKVKKQIF 177
 D+P+Y GT++ L KG GHLEGTSLFVG T SV+T HRGL A +P+L+KVK G
30 Sb|ct: 99 DLFVYHGTAADDLLKGLGHLEGTSLFVGGEGRSVITGHRGLAEATMFINLHKVKITQDSL 158

Query: 178 YVINIKETLAYKVVSIKVVDPITALSEVKIVMGKDYITLLTCTPFMINSRHLNVRGERI 235
 V E L Y+V S KVV+P +++ GRD +TL+TCTP IN+HR+L+ GERI
30 Sb|ct: 159 IVEVFGEVLYFRVTSIKVVEPEETEARLVEBGRDLLTLVTCTPLGINTHRILLTGRI 216

```

There is also homology to SEQ ID 3740.

35 SEQ ID 3916 (GBS208) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 5; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 8; MW 59.7kDa) and in Figure 160 (lane 5; MW 60kDa).

GBS208-GST was purified as shown in Figure 224, lane 7-8.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1272

A DNA sequence (GBSx1349) was identified in *S.agalactiae* <SEQ ID 3917> which encodes the amino acid sequence <SEQ ID 3918>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.13 Transmembrane 265 - 281 (260 - 284)
50 ----- Final Results -----

```

-1409-

bacterial membrane --- Certainty=0.4652 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP:ARC13546 GB:AP019629 putative fimbria-associated protein  
 [Actinomyces naeslundii]  
 Identities = 96/265 (36%), Positives = 150/265 (56%), Gaps = 10/265 (3%)

10 Query: 41 QASHANINAFKEAVTKIDRVETNRRLAYAYNAGI-AGAKTNGEYPALKDPYASBQKA 99  
 Q + + + A A E + ++E A+YN ++ AGA P A +  
 Sbjct: 15 QYNQSKVTADYSAQVDGARPDAKTQVBQAHYNDALAGAVLEANNHV---PTGAGSSKD 71

15 Query: 100 GVVEYARMLEVKIQ--IGHVIIPRINQDIPTIYAGSABENLQRGVCHLECTSLPVGSETH 157  
 ++YA +L+ + + + IP I+ D+P+Y G+H++ L +G+CHLECTSLPVGGE T  
 Sbjct: 72 SSLQYANILKANNBGLMARLKITPSISLDFVYHGTADDTLLKGLCHLECTSLPVGSETR 131

Query: 158 AVLTAHRLGLPTAKLFTNLKDVTVGRDFYIEHIGKIAVQVQI KVIAPQIREDLYVIQSE 217  
 +V+T HRLG A +FTNLKRV GD +E G +Y+V KV+ P++ E L V +G+  
 20 Sbjct: 132 SVITGHRGLAETMFTNLKVKVTDGLIVEVFGEVLTIRVTSTKVEPEETEARVEEGK 191

Query: 218 DRVTLTCTPYMINSHRLLVGRKRI-PYVEKTVQKDSKTFRQQQLTYAMMVVGLILLS 276  
 D +TL+TCTP IN+HR+L+ G+RI P K + + + + ++ + GLI+ +  
 25 Sbjct: 192 DLLTLVTCTPLGINTHRIILLTGERIYPTPAKDLAAGKRDPVPHFPMWAGLAAGLIVVG 251

Query: 277 LLIM---FKKTKQKGRKNEKAAQ 298  
 L +W + + K+R A+Q  
 Sbjct: 252 LYLMBRGYAAARAKERALARAKAAQ 276

- 30 There is also homology to SEQ ID 3740.

SEQ ID 3918 (GBS209) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 4; MW 62kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 3; MW 37.2kDa).

GBS209-His was purified as shown in Figure 221, lane 8.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1273

A DNA sequence (GBSx1350) was identified in *S. agalactiae* <SEQ ID 3919> which encodes the amino acid sequence <SEQ ID 3920>. Analysis of this protein sequence reveals the following:

40 Possible site: 27  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -9.66 Transmembrane 281 - 297 ( 276 - 300)

----- Final Results -----

45 bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:BAE04030 GB:AP001508 unknown [Bacillus halodurans]  
 Identities = 45/141 (31%), Positives = 63/141 (43%), Gaps = 20/141 (14%)

Query: 153 TQRLDLLKVGVDGDKTKPLAGVVFELYEKGRTPIRVKNVHSQDIDAARHLETDSSGHI 212  
 TG L++ KV D DT + L G F LY+ G IR LFT G  
 55 Sbjct: 1084 TGSLEVTKV--DADTGEVLQGATFTLYDSGEFAITPT-----LETGEDGKA 1127

Query: 213 RISGLIHGDYVLKEIETQSGYIGQAEAVTIEKSKTIVTIENKKVPTPKVPSRGL-I 271  
 L++GDY+LKE GY +G +T + VT+EN+K +V + G + +  
 Sbjct: 1128 TFWNLLYGDYLLKESDAPGYLVGINDTQRTVIDTVLHEVTIENKESDINRVSAGVAVQL 1187

Query: 272 PKTGEQQAMALVIGIILIAL 292  
 K E+ +L G L AL  
 Sbjct: 1188 QKVDEETGESI---QGALFAL 1205  
 Identities = 64/259 (24%), Positives = 113/259 (42%), Gaps = 48/259 (18%)

Query: 16 GTMFGISQT--VLAQETHOLTIVHLEARDIDRPNP---QLRIAPKE-GTPIEGVLYQL 67  
 G + GI+ T + H++T+ + E DI+R + QL+ +E G ++G L+ L  
 Sbjct: 1147 GYLVGINDTQRTVIDTVLHEVTIEN- EKSDINRVSAGVAVQLQKVDEETGESLQALFAL 1205

Query: 68 YQLKSTEDGDLAHWNLSLTITELKKQAOQVFATTNOCGRATFNOLPDGIYYGL----AV 123  
 Q E +TI E++ + + A + + G F + L + Y L V  
 Sbjct: 1206 QQKVDDE-----FVTIAGMETDEEGIVFAGSLIEFGDYQFVELNAPVGVKLDETPVV 1256

Query: 124 KAGEKNENVSAPLWDLSEDKVIYPKIINSTGLDLKLVGVGDGTPKFLAGVVFELYKNG 183  
 E+R+ + ++L ++ + P G + L+K V D D L G F L + G  
 Sbjct: 1257 FTVEEDRTET---TELQKENHLIP-----GSVQLVKVDAD-DAANTLGAETLLDQGG 1306

Query: 184 RTPIRVNGVHSGPDIDAAKHLETDSSGHIRISGLIHGDYVLKEIETQSGYIGQAEAVT 243  
 V+ G L TD +G + + L G+Y E + +GY+ T  
 Sbjct: 1307 NV---VQEG-----LTTDENGQVVVITLKPGEYQFVETKAPAGVELEATPIGT 1352

Query: 244 IEKS--KTVTVTIENKKVP 260  
 IE++ + TV +EN +P  
 Sbjct: 1353 IERNQGEVATVAENHLIP 1371

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3920 (GBS52) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 4; MW 30.5kDa).

GBS52-His was purified as shown in Figure 192, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1274

A DNA sequence (GBSx1351) was identified in *S.agalactiae* <SEQ ID 3921> which encodes the amino acid sequence <SEQ ID 3922>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -6.26 Transmembrane 554 - 570 ( 551 - 575)  
 INTEGRAL Likelihood = -0.16 Transmembrane 34 - 50 ( 34 - 50)

----- Final Results -----  
 bacterial membrane --- Certainty=0.3506 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8779> which encodes amino acid sequence <SEQ ID 8780> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0  
 McG: Discrim Score: -5.81  
 GvH: Signal Score (-7.5): -1.92  
 Possible site: 37  
 >>> Seems to have a cleavable N-terminal signal sequence  
 ALOM program count: 2 value: -6.26 threshold: 0.0



-1411-

INTEGRAL Likelihood = -6.26 Transmembrane 527 - 543 (524 - 548)  
 PERIPHERAL Likelihood = 5.36 194  
 modified ALAM score: 1.75

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3506 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

LEXTG motif: 521-525

The protein has homology with the following sequences in the GENPEPT database.

15 >GP\_CAA57459 GB:X81869 crf2 [Lactobacillus leichmannii]  
 Identities = 140/505 (27%), Positives = 220/505 (42%), Gaps = 94/505 (18%)  
 Query: 102 GEVISNYAKLGNVKEGLQGVQPKRYKVTIDI-----SVDELKKLITVEAADAKVGTILEE 156  
 GE++++ G L GV FK Y V S D + T +DAK L  
 20 Sbjct: 58 GEINNDPFGGTG-----LNGVTFKAYIVTDHYLSLRKSGDSQAQVATQISDAKDSNLP 112  
 Query: 157 --GVSLPQKTNAGLVVDAL-----DSKSNVR- YLYVEDLKNPSNITKAYAVFFV 204  
 G ++ +T A D + DS N +YL+VE +SP+++T A P V  
 25 Sbjct: 113 YAGSAIATSTTATSKGEGDIAAFDNLNKDSGNGYQTYLFVET--DSPDTVTQQ-AAPIV 169  
 Query: 205 LELPVANSTGTGFLS-EINIYPKNVVTDPEKTDKDVKKLQGDAGYTI-----G 252  
 L +P+ ++ T ++ +I IYPKNV + P T KD+ + + D T+ G  
 Sbjct: 170 LTMPIYKTSQTSANRHDIIQIYPMKVKST-PIT-KLDEASKKDLAVTLDPGSGTINYAQYG 227  
 30 Query: 253 EEFKFLKSTIPANGLDYEFKITDKFADGLITYSGKIKSGKSTINARDEHYITDEPTVD 312  
 + F + + +P N+ D + F + DK G+ + + L + YT+++  
 Sbjct: 228 KSPGYNITVNVPMNKKDKDTPNVVDKPTGI--DIDASTVSDGLKSTDIYVKN--- 280  
 35 Query: 313 NQNTLKITIKPKPKELIAELLKGMTLVKNQDALQATANTDAAFLPIVASTINEKAVL 372  
 N ++ FK + L G +L I ++T A  
 Sbjct: 281 KDNGYQVVKFTS--AAVQALAGKSLT-----ITYKATLTNNATP 318  
 Query: 373 GKAIENITFLQYDHTPOKADNPKPSNPPRKPEVHTGGKRFVKKDSSTETQTLQGAEDFLA 432  
 KAI NT L + + S P P ++TGG +FVKDS +TL GAEP L+  
 40 Sbjct: 319 DKAIENITATLSIGNGTINIT-----STPANGPRIYTGGAQFVKKDSQNKSTLAGAEFLVK 373  
 Query: 433 --SDGTAVKMTDALIKANINOFYIAGEAVTGQPIKLKSHDTGTFEIKGLAYADANABST 490  
 S+G V + + N A EA T S +G +KGL+Y ++ +  
 45 Sbjct: 374 VDSNGNIVSYATQASDGSYTWDSATEATT-----YTSANGLVALGLSY---SDKLDS 425  
 Query: 491 AVTYLKKETKAPESYIVDPKIEFTVSQTSYNTKPTDITVDSADATPDITKNNKPSIPN 550  
 +Y L E +AP+GY D ++F+++Q S+ D+ TI N K +P+  
 Sbjct: 426 GESYALLEIQAPDQYAKLDSPVKFSITQGSF-----GDSNKITIDNTEKGLLS 474  
 50 Query: 551 TGGIGTAIPVIGAAVMAFAVGMK 575  
 TGG G IF+IG +M A GK  
 Sbjct: 475 TGGGIIYIFLAIGVYIMIVARGGYK 499

No corresponding DNA sequence was identified in *S.pyogenes*.

55 SEQ ID 8780 (GBS80) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 6; MW 56.8kDa).

The GBS80-His fusion product was purified (Figure 104A; see also Figure 194, lane 5) and used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 104B), FACS (Figure 104C), and in the *in vivo* passive protection assay (Table III). These tests confirm  
 60 that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1275

A DNA sequence (GBSx1352) was identified in *S.agalactiae* <SEQ ID 3923> which encodes the amino acid sequence <SEQ ID 3924>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4043 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1276

A DNA sequence (GBSx1353) was identified in *S.agalactiae* <SEQ ID 3925> which encodes the amino acid sequence <SEQ ID 3926>. This protein is predicted to be MsmR. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.01 Transmembrane 75 - 91 (75 - 92)

----- Final Results -----
bacterial membrane --- Certainty=0.1404 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9679> which encodes amino acid sequence <SEQ ID 9680> was also identified.

SEQ ID 3926 (GBS360) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 9; MW 74kDa).

GBS360-GST was purified as shown in Figure 216, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1277

A DNA sequence (GBSx1354) was identified in *S.agalactiae* <SEQ ID 3927> which encodes the amino acid sequence <SEQ ID 3928>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1762 (Affirmative) < succ>
```

-1413-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3929> which encodes the amino acid sequence <SEQ ID 3930>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1640 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 15 Identities = 93/98 (94%), Positives = 96/98 (97%)  
 Query: 1 MDKIIKSIASGAPRSYVLDSTETVKLAQEKHLLSSSTVALGRTLIANQILAAKQKGS 60  
 MDKIIKSI+ SGAFR+YVLDSTETV LAQEKH+TLSSSTVALGRTLIANQILAAKQKGS  
 20 Sbjct: 1 MDKIIKSIAGSAPRAYVLDSTETVALAQEKHNTLSSSTVALGRTLIANQILAAKQKGS 60  
 Query: 61 KITVKVIGDSSFGHIIISVADTKGHVKGVIQNTGVDIKK 98  
 KITVKVIGDSSFGHIIISVADTKGHVKGVIQNTGVDIKK  
 Sbjct: 61 KITVKVIGDSSFGHIIISVADTKGHVKGVIQNTGVDIKK 98

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1278

A DNA sequence (GBSx1355) was identified in *S.agalactiae* <SEQ ID 3931> which encodes the amino acid sequence <SEQ ID 3932>. Analysis of this protein sequence reveals the following:

- 30 Possible site: 17  
 >>> Seems to have an uncleavable N-term signal seq  
 ----- Final Results -----  
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 40 >GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]  
 Identities = 34/48 (70%), Positives = 39/48 (80%)  
 Query: 1 MQEVLIIARENHQVTHEHVSILLTCVQRELIVEVNTQPLSREPREKYM 48  
 + EV IIA+ NHQVTHEHVSILLTC+QRLI EV +T PLS +F KYM  
 45 Sbjct: 70 VHEVFIIAKTNHQVTHEHVSILLTCIQELIKEVEKTPLSRSDPCMKM 117

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1279

- 50 A DNA sequence (GBSx1356) was identified in *S.agalactiae* <SEQ ID 3933> which encodes the amino acid sequence <SEQ ID 3934>. This protein is predicted to be TnpA (orfB). Analysis of this protein sequence reveals the following:

-1414-

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.5248 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9907> which encodes amino acid sequence <SEQ ID 9908>  
 10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 9677> which encodes amino acid  
 sequence <SEQ ID 9678> was also identified. A further related GBS nucleic acid sequence <SEQ ID  
 10911> which encodes amino acid sequence <SEQ ID 10912> was also identified.

There is homology to SEQ ID 1336.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 15 vaccines or diagnostics.

#### Example 1280

A DNA sequence (GBSx1357) was identified in *S.agalactiae* <SEQ ID 3935> which encodes the amino  
 acid sequence <SEQ ID 3936>. Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25                   bacterial cytoplasm --- Certainty=0.4489 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB64982 GB:U43834 Ydr540cp [Saccharomyces cerevisiae]  
 Identities = 93/171 (54%), Positives = 121/171 (70%), Gaps = 3/171 (1%)

30                   Query: 1   MRVYENKSELKKEISKTFPEKYIMEFNINPEHLKDKRIDEVDRTPAANLSYQVQWTHLVK 60  
                   MR Y +K+ELK+RI K +EKY EF I E+ KD++++ VDRTP+ NLSYQ+GM NL+L+  
                   Sbjct: 1   MREYTSKKELKEEIEKKYKYDAEPETISQKQKDEKVTVDRTPSHNSYQIGWVTHLLE 60

35                   Query: 61   WEEDERKGLQVTPSDKFGWVGLGELYQWFTDITYAHLSELQELKAKLANENINSIYAMIDLL 120  
                   WE E G V+TP+ +QWN LG LKQ F Y S++E +AKL E +N +Y I L  
                   Sbjct: 61   WEAKEIAGYNVETPAPGYQWNLGSLYQSFYKCYGIYSIKEQRAKLREAVNEVTHISTL 120

40                   Query: 121   SEELPFAHMRKMGADATKATWVYKFIHVNIVAFPTPRTKIRKWKKIV 171  
                   S++ELF+A RHW AT A N VYK+IH-NTVAFP PR KIRKWK++V  
                   Sbjct: 121   SDDELPGAGNRKW---ATTQWMPVYKWIHNTVAFPTNFRGKIRKWKRLV 168

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 45 vaccines or diagnostics.

#### Example 1281

A DNA sequence (GBSx1358) was identified in *S.agalactiae* <SEQ ID 3937> which encodes the amino  
 acid sequence <SEQ ID 3938>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have an uncleavable N-term signal seq  
           Integral   Likelihood = -3.45   Transmembrane   10 - 26 ( 2 - 26)

50

-1415-

## ----- Final Results -----

bacterial membrane --- Certainty=0.2381(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8781> which encodes amino acid sequence <SEQ ID 8782> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6  
 MoG: Discrim Score: 8.80  
 GVH: Signal Score (-7.5): -3.94  
 Possible site: 28  
 >>> Seems to have an uncleavable N-term signal seq  
 ALCM program count: 1 value: -3.45 threshold: 0.0  
 INTEGRAL Likelihood = -3.45 Transmembrane 7 - 23 ( 2 - 26)  
 PERIPHERAL Likelihood = 10.40 69  
 modified ALCM score: 1.19

## \*\*\* Reasoning Step: 3

## ----- Final Results -----

bacterial membrane --- Certainty=0.2381(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68889 GB:Y07615 acid phosphatase [Haemophilus influenzae]  
 Identities = 112/245 (45%), Positives = 148/245 (59%), Gaps = 10/245 (4%)  
 Query: 5 MKKVLVSSLLVLGITITLQTVVEAKGPKVAYTQSGHTALSDTNKDKVTTISIDRIQKSL 64  
 MK V+ S++ L +T V G YTG G A + + IS+D+I++SLS  
 Sbjct: 1 MKNVMKLSVIAL---LTAALVEMAKGTETPTTQSGTNAREMLQEQAIHWISVDQIKQSL 57  
 Query: 65 GKPKITVSPFDIDDTLLFSSQYFQYQKEYVTGSPDFLHKQKFWDLVAKRGDQDSIPKEYA 124  
 GK PI VSPFIDDT+LFSS F+G++ +PG D+L Q FW+ V D+ SIPK+ A  
 Sbjct: 58 GKAPINVSFDIDTVLFSSPCFYHGQCKFSPGKHDLKNGDFWNEVWAGCDKYSIPKQIA 117  
 Query: 125 KKLIAHQKRGDKIVFITGRTRGSMYKGEVDKTKAKLAKDFELDKPIAVNYTGDKPKK 184  
 LI MIQ RGD++ F TGR T G+VD L K F + V + G + ++  
 Sbjct: 118 IDLINMHQARGDQVYFTTGR T-----AGKVDGVTPILSKTFNIKMHFVPEMGR-ERT 170  
 Query: 185 YKYDKSYIYIKYGSDDIHYGSDDDINAREAGARPILRAPNSTNLPLPAGGYSSEVL 244  
 KY+K+ I + IHYGSDDD+ AA+EAG R IR++RA NST P+P GSGYSEVL  
 Sbjct: 171 TKYNKTPALISHKVS IHYGSDDDVLAKEAGVRSIRLMAAANSTYQPMPTLGGYSEVL 230  
 Query: 245 ENSAY 249  
 NS+Y  
 Sbjct: 231 INSSY 235

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3939> which encodes the amino acid sequence <SEQ ID 3940>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -3.98 Transmembrane 6 - 22 ( 4 - 25)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2593(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA68889 GB:Y07615 acid phosphatase [Haemophilus influenzae]

-1416-

Identities = 105/237 (44%), Positives = 141/237 (59%), Gaps = 10/237 (4%)

Query: 9 LFTVSPGCIILALPVASGPKVPTQEGITA--ISNQATVKLISADIASSLEQKPKITVS 66  
 L ++ A+P A G P T Q G A + + + IS+ I SLGG+ P I VS  
 5 Sbjct: 7 LSVIALLTAAAVPAMA-GKTEPTQSGINAREMLQBQAIHWISVDGIKSLGKAPINVS 65

Query: 67 FDIDDITLLFTSQYFQYQKEYITPGSDFLHKQKFWDLVAKRGDQDSIPKEYAKQLIAMHQ 126  
 FDIDDT+LF+S F +G++ +PG D+L Q FW+ V D+ SIPK+ A LI MHQ  
 10 Sbjct: 66 FDIDDTVLFSSPCFYHQQKPSGKHIDYLNQDFWNEVNAGCDKYSIPKQIAIDLINMHQ 125

Query: 127 KRGGKIVFITGRTRGSMYKKEIDKTAKSLAKDFLKDPIAINYNTGDKAVKPYQYDKTY 186  
 RGD+ F TGR T G+D L K F + + + G + + +Y+KT  
 Sbjct: 126 ARGDQVYFTTGR-----AGKVDGVPILEKTFNKNMHPVFVMSRE-RTTKYNKTTA 178

Query: 187 IKKQSGSIHYGDSDEDINAKEAGARPRIILRAPNSTNLPLPKAGGYCEEVLNSAY 243  
 I + IHYGDS+D+ AAKEAG R IR+RA NST P+P GGYCEEVL NS+Y  
 15 Sbjct: 179 IISHKVSIIHYGDSDDVLAKEAGVGRGIRLMFAANSTYQMPITLGGYCEEVLINSSY 235

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 196/245 (80%), Positives = 216/245 (88%), Gaps = 2/245 (0%)

Query: 5 MKKVLSLLVLGITITLQTVVEAKGPKVAYTQEGMTALSETNKKVTTISIDBIQKSL 64  
 MKK S L + + VEA GK V Y T Q E G + T A + S N + V I S I + I S L E  
 25 Sbjct: 1 MKKEFTSILFTVSPGCIILALPVASGPKVPTQEGITALS--NQATVKLISADIASSLE 58

Query: 65 GKPKITVSFDIDDTLLFSQYFQYQKEYITPGSDFLHKQKFWDLVAKRGDQDSIPKEYA 124  
 G+KPI TVSF D I D D T L L F + S Q Y F Q Y G K E Y + T P G S D F L H K Q K F W D L V A K R G D Q D S I P K E Y A  
 30 Sbjct: 59 GKPKITVSFDIDDTLLFTSQYFQYQKEYITPGSDFLHKQKFWDLVAKRGDQDSIPKEYA 118

Query: 125 KQLIAMHQKRGDKIVFITGRTRGSMYKKEVDKTAALAKDFLKDPIAINYNTGDKPKP 184  
 K+L I A M H Q K R G D K I V F I T G R T R G S M Y K + G E + D K T A K + L A K D F L K D K P I A I N Y T G D K K P  
 35 Sbjct: 119 KQLIAMHQKRGDKIVFITGRTRGSMYKKEIDKTAALAKDFLKDPIAINYNTGDKAVK 178

Query: 185 YKYDKSYIYIKKQSGSIHYGDSDDIHAAREAGARPRIILRAPNSTNLPLPEAGGYCEEVL 244  
 Y+YDK+YYIKK GS IHYGDS+DI+AA+EAAGARPRIILRAPNSTNLPLPEAGGYCEEVL  
 40 Sbjct: 179 YQYDKTYIYIKKQSGSIHYGDSDEDINAKEAGARPRIILRAPNSTNLPLPKAGGYCEEVL 238

Query: 245 ENSAY 249  
 ENSAY  
 40 Sbjct: 239 ENSAY 243

SEQ ID 8782 (GBS100) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 5; MW 28kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 2; MW 53kDa).

45 The GBS100-GST fusion product was purified (Figure 106A; see also Figure 197, lane 4) and used to immunise mice (lane 1 product; 9.9µg/mouse). The resulting antiserum was used for Western blot (Figure 106B), FACS (Figure 106C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1282

A DNA sequence (GBSx1359) was identified in *S.agalactiae* <SEQ ID 3941> which encodes the amino acid sequence <SEQ ID 3942>. Analysis of this protein sequence reveals the following:

55 Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

-1417-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3288(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 **Example 1283**

A DNA sequence (GBSx1360) was identified in *S.agalactiae* <SEQ ID 3943> which encodes the amino acid sequence <SEQ ID 3944>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4004(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9675> which encodes amino acid sequence <SEQ ID 9676> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25

>GP:BAE04406 GB:AP001509 RNA methyltransferase [Bacillus halodurans]  
 Identities = 199/452 (43%), Positives = 300/452 (65%)

Query: 12 KRKIMLHKNDIETETSDISHEGMOIAKVDGCVFFVENALPGEIIRMRVLKLRKRGYK 71  
 K++ ++KND++E I D++H+G G+AKVDG+ F+ ALPGE +K +V+K++K G+G+  
 Sbjct: 3 KQAPVNRKNDVVEVTIEDLHDGAGVAKVDGYALFIPKALPGBERLKAIVKVKKGYPGR 62

30

Query: 72 VEEYLTTSPHNEGLDYTLRTGIADLGHLYEQQLLFKQKQVADNLKYIAHISDLVLEP 131  
 V ++ SP R E ++ G L H++Y+ QL +KQKQV D L +I I+ V V P  
 Sbjct: 63 VLNIEASPDRVEAPCPVFNQCGGQQLQHMSYDAQLRYKQKQVDLIERIGKITAVTVRP 122

35

Query: 132 TLGWTIPLAYRNKAQVPVRRVDQQLTGFRRINSFTLVSIEDYLIQEKEIDALINFTRDL 191  
 T+GM P YRNKAQVPV +G L GF++ SH ++ ++ +IQ +E D +I ++L  
 Sbjct: 123 TIGNEPWRMYRNKAQVPVGRERGGGLAGFYQERSERIIDMECMIQHEENDKVIQGVKL 182

40

Query: 192 LRKFDVKPYDEBQOGLIRNLVVRGHYTGQLMLVLTTRPKIFRIDQMIEKLVSAPFSV 251  
 R+ ++ YDEE+ G +R++V R G TG++M+VL+T ++ +IE++ A P V  
 Sbjct: 183 ARELGIRGYDEEKHRTLRHVVARVYKGTGRIMVVLITRGESLPHKKTFLIERIKHAI 242

45

Query: 252 VSIQNNINURNNSNVIHGKFRFYLGSUTIEDQMLGNITAIASQSFYQNNYEMARKLYQKA 311  
 SI+QNN++ +NVIFG ++ L+G + I D + -AISA+SPYQVN E + LY +A  
 Sbjct: 243 KSIVQNNPKRTNVI GDKTKVLMGSEYTYDITIGDKIPAIARSFYQNNPEQTKVLVYDQ 302

50

Query: 312 IDPSDLNSEDIVIDAYSIGITIGLSVAKQVHVGVVEVKA/SDAKENATRNIGTINSTY 371  
 ++F+L + VIDAY GIGTI L +A+Q KHVYGV+V +A+SDAK NA NG N +  
 Sbjct: 303 LEFANVTGSETVIDAYCIGTISLFLAQQAKHVYGVKIVPFAISDAKRNARLNGFANVQ 362

55

Query: 372 VADSAENAMAKLKEGIKPTIVMVDPPKRLTSPFVYSAQTKADKITYISCNATWARD 431  
 AE M W +G++ VI+VDPFRKG E+ + + K D++ Y+SCN AT+ARD  
 Sbjct: 363 AVGDARKVPMWYAGVRAVIDVDPKPKQDEALLLKTLIMKPEPVRVYSCNPATLARD 422

Query: 432 IKLFELGYHLVKIQPVOLFPMTHHVCVALL 463  
 +++ E+ GY +QFVD+FP T H+E VA+L  
 Sbjct: 423 LRVLBDGYETKDVPQVDMFPWTHHSVAVL 454

-1418-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3945> which encodes the amino acid sequence <SEQ ID 3946>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1262 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 332/454 (73%), Positives = 387/454 (85%)

Query: 12 KRKIMLRKNDIIEISDISHEGMGIKVDGFVFVFNALPGEIIMRVLKLKRKIGYVK 71
KR ML KNDII+ ISD+SHBG G+AK DGFVFFV+HALP E+I MRVLK+ K G+GK
Sbjct: 8 KRIRMLKGRNDIIQVAISDLSEHGAGVAKHDGFVFVDNALPREEVDMLKVNKNSGPGK 67

Query: 72 VEEYLITSPHRNGBGLDLYTLRTGIADLGHLYEQQLLFKQKQVADNLYKIAHISDVLVEP 131
VE Y S RN ++ TYLRTGIADLGHLYE QL FK+KQV D+LYKIA ISDV VE
Sbjct: 68 VEAYHYLSARNADVHLTYLRTGIADLGHLYEQLTFKKKQVQDSLYKIAGISDVTVES 127

Query: 132 TLGNTIPLAYRNKAQVPRVRDQGLTGFRRKNSHTLVISDYLIQKEIDALINPTDRL 191
T+GMT PLAYRNKAQVPRVR+QGLETFGFRK+SH L+ I DY IQ+KEID LINPTDRL
Sbjct: 128 TIGMTEPLAYRNKAQVPRVRNGQLETFGFRKSHDLIPISDYIQQEIDRLINPTDRL 187

Query: 192 LRKFDVKPYDESSQSGSLIRNLVVRGHYTGQMLVLVITTRPKIFRIDQMEIKLVSAFSPV 251
LR+FD+KPYDE +Q+GL+RN+VVRGHY+G++MLVLVITTRPK+FR+DQ+ISK+V AP+V
Sbjct: 188 LRKFDVKPYDETEQTLGRNLNVVRGHYSGEMMLVLVITTRPKVFRVDQVIEKIVRAPEAV 247

Query: 252 VSIQNINDRNSNVIPGKEFTLYGSDTIEDQMLGNTYATSAQSIFYQNTMAEKLYQGA 311
VSI+QNIND+N+N IPGK+F+TLYG DTI D MLGN YATSAQSIFYQNT MAEKLYQ
Sbjct: 248 VSIQNINDKNTNAIPGDKFTLYGKDTITDSMLGNNTYATSAQSIFYQNTVMAEKLYQTA 307

Query: 312 I FSDLSNSEDIVIDAYSIGITGLSVAQKVHVHVGVEVKAQSDAKENATNGSITNSTY 371
I FSDL+ +DIVIDAYSIGITGLS AK VK VYGVEV+E AV DA++NA NGSITN+ +
Sbjct: 308 IAFSDLSKDDIVIDAYSIGITGLSFAKTVKAVYGVEVIEAARVDAQQAALNGSITNAV 367

Query: 372 VADSAENAMAKLKEGIKPTVMVDPPRKGLTESFVYSAAQTAKDKITYISNSATWARD 431
VAD+AE+AMA W K+GIKP+VI+VDPPRKGLTESF+ ++ KITY+SCN ATWARD
Sbjct: 368 VADTAENAMATWAKDGKPSVILVDPPRKGLTESFQASVAMGPQKITVYSCNPATWARD 427

Query: 432 IKLFEELGYHLVKIQPVDLFPWTHRVCEVALLVK 465
IK ++ELGY L K+QPVDFP THHVECV LL+K
Sbjct: 428 IKRYQELGYKLAKVQPVDFLPQTHHVECVLLIK 461
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1284

A DNA sequence (GBSx1361) was identified in *S.agalactiae* <SEQ ID 3947> which encodes the amino acid sequence <SEQ ID 3948>. This protein is predicted to be PSR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood =-12.15 Transmembrane 135 - 151 (127 - 155)

----- Final Results -----
bacterial membrane --- Certainty=0.5861 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```



The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB76822 GB: AJ276232 PGR protein [Enterococcus faecalis]  
Identities = 143/409 (34%), Positives = 206/409 (49%), Gaps = 56/409 (13%)

5 Query: 48 QRRTSPP--TNSYYERPYSDEYYQDDDFYSRPQLTSQQLPIYQERAPKKKKQARKEK 105  
+ R E P S E Y DSY +D T G ++ P+ KK + K+K  
Sbjct: 31 EHRERPEBELAESLQEPYEDSYTDSRRSERRHCTDSGG-NGSDQPPRGKKDKKPKKK 89

10 Query: 106 QRVKMAPPPKAITPPRKKKKKGLKFIGILLIVLGMVFMFKMGRVNNKGSHYS 165  
RKK K K F K++ I+L+++ MF+NG + S  
Sbjct: 90 -----RKKSKTKRFFKWLIVLILLPAYSITVMFLKKGSAAEHDS-LE 131

15 Query: 166 PAIEDFKGKDAVDGT-NILLGSDKRVSRSTDAKDTIMVNVGNKDNKVMVSPWRD 224  
+E F G + +G NILLGSD R + R DTIMV + K K+SPWRD  
Sbjct: 132 QKCVSTFNGVSSNGARNILLGSDTRGEDAG---RADTIMVLQNGPSKKPKLISFWRD 188

20 Query: 225 LLVNIPIYSTEGYDMKLNASFNLGBQDNHKGAEYVRQTLKNHFDIDIKTYVMVDFETFA 284  
V+IP G K+NA++ G GAE VR+TLK +F+D KY Y VDF+P  
Sbjct: 189 TFDVIP----GVGPNKINAYAYG-----GAEIVRETLKQNFNLDTKYAKVDQSPFE 237

25 Query: 285 DAIDTLFPNGVKINAKFGLVGGQSDSVKVPDRLRMQNGVVPQKIKVGIQYMDGRTLIN 344  
+D++FP GVKI+A+ L +D V I+ G Q MDG LL  
Sbjct: 238 KIVDSMFPGVKIDAEKSL-----NLGDVD-----IEKQQVMGHVLLQ 277

25 Query: 345 YARFRKDDGDFGRTORQQQVMRAIVSQIKDPRLFTGSAALGQYALTSSNLGYSFVL 404  
YARFR D++GDFGR +RQQQVM A++SQ+K+P L ++GK S+++ SFALT  
Sbjct: 278 YARFRMDGDFGRVRRQQQVMRAVSMKMNPMILLETPESLGLKVMGNTDVPVPSFVL 337

30 Query: 405 DGIPILSDARNGIKQMTIPREGDWDVDDYQYQQGLTIDFAKYKKILK 453  
+G +L K G++ +++P W Y G L +D K ++K  
Sbjct: 338 NGPSLLIKGKTGVESLVFPVDSNPGESSYAGSILEVDEGKNADAEK 386

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3949> which encodes the amino acid sequence <SEQ ID 3950>. Analysis of this protein sequence reveals the following:

35 Possible site: 49

>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -7.96 Transmembrane 159 - 175 ( 152 - 180)

40 ----- Final Results -----  
bacterial membrane --- Certainty=0.4185 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

>GP: CAB76822 GB: AJ276232 PGR protein [Enterococcus faecalis]  
Identities = 140/345 (40%), Positives = 195/345 (55%), Gaps = 41/345 (11%)

50 Query: 140 PRSQK---RRKKKKGCKMKWFFNLLGLLMTVLNGLGFMFAKQVFDISTNKANYKPAVSQ 195  
PR +K +K +KK K FF L +LL+ + +MF KG + + + V +  
Sbjct: 78 PRGKKKKKKKKKKKKSKTKRFFKWLIVLILLPAYSITVMFLKKGSAAEHDSLPQCKV-R 136

55 Query: 196 AFDQGETDQGT-NILLGSDQRVTQSTDAKDTIMVNVGNHAKIKRMVSPMRDTLINI 254  
F+G ++ +G NILLGSD T+G R DTIMV + +KK K++SPMRDT ++I  
Sbjct: 137 TFGVKSNGAKNILLGSD---TRGEDAGRADTIMVLQNGPSKKPKLISFMRDTFVDI 193

60 Query: 255 PGYSYNDNSYDLKLNASFNLGBQEDHKGAEYVRRALKHNFIDIDIKTYVMVDFETFAEAD 314  
PG N K+N+A+ G GAE VR LK NF++D KY Y VDF+P + +D  
Sbjct: 194 PGVGFN-----KINAYAYG-----GAEIVRETLKQNFNLDTKYAKVDQSPFEKIVD 241

Query: 315 TLPFNGVKIDAKPATVGVAVDSVVPDRLRMQNGVVPQKIEVGBQRNDGRTLLIYARF 374  
++FP GVKIDA+ + + +D V+ IE G+Q MDG LL YARF  
Sbjct: 242 SMFPGVKIDAEKSL-----LNLGDVD-----IEKQQVMGHVLLQYARF 281

-1420-

Query: 375 RKDDRGDPRTVRQQQVMSAVMSQIKDPTKLTGSAAGIKIYALTSTNVSPFPVVKNGVS 434  
 R D+RGDPR RQQQVMSAVMSQ+K+P L ++GK+ ST+V F++ NG S  
 Sbjct: 282 RMDRGDPRTVRQQQVMSAVMSQMNPMILRTPESLGKLGVMETDVPVSPMLTNGPS 341

5 Query: 435 VLGGKNGVHVHTIPENGWDVDEIMYGGQALYIDFKYQKTLAK 479  
 +L GK GVE +++P W Y G L +D K + K  
 Sbjct: 342 LLIKGTGVESLSVPVPDSWMPGESSYAGSLLEVDEQKNADAIEK 386

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 273/486 (56%), Positives = 340/486 (69%), Gaps = 32/486 (6%)

Query: 1 MSRNTYQQQLNHHEELRYNYLLNHYLNHREKMEQFYLHYKKTAVRQRRTESPFTNSGY 60  
 N++ G L+HHEELRY YLL+N+ YL+E EK EF +L K R ++ S  
 Sbjct: 1 MTKYPMGGLSHHEELRYFYLLNLSYLSNKKKEFAPLKSLEIGRAYAPSQHYRKSQR 60

15 Query: 61 EEPY-SDSY- - - - - QDDDFYSEPOLTSQGLPIYQERAPFKKKQARKEKQKVY 110  
 +EPY D YY +DD + GLPIY +E KK K R +  
 Sbjct: 61 QSPYFSEDYNDYSNDILLEDVDVNDSSVFYGLPIYPKEDRYLNKKT-- - KLTARRPI 117

20 Query: 111 NADPE- - - - - PKAITPPRKKK- - - - - FKGLFKFGLILLIYVLSGMVPMVK 152  
 AD P P++ KKK K F +G++L+ VL G+ HF K  
 Sbjct: 118 DAPQPIDEDDAFLTSVARCALPRSQKRKHKKGKGMWFPNILLIYVLSGLMFAK 177

25 Query: 153 GMRDNGKSHYSPIAIEDFGKGDVQDGNILILGSDKRVSESTARTDTIMVAVNGK 212  
 G+ D++ K++Y PA+ + F G++ DGNILILGSD+RV++ STARDTDMV NVGN  
 Sbjct: 178 GVFDISTNKANYKPAVSQAQDQGTQDGNILILGSDQRTVQSTDARTDTIMVAVNGH 237

30 Query: 213 DNKVMVSMFRDLLVNIENYS-TEGYVMKLAASPHLGEQDNHGAERYVRLTKNHFD 271  
 K+KMSVFNRL L+NIP YS + YD+KLN++PHLGEQ++H GAERYR+ LK++FQID  
 Sbjct: 238 AKKIMVSMFRDTLINIPGYSTNENSYDLKLSAFLHGEQEDHNGAERYRRLALGNHFD 297

35 Query: 272 IKYYVMVDFETFAAIDTLFPNGVKINAKFGLVGGQSADSQKVPDILRMKNGVVPQIK 331  
 IKYYVMVDFETFAAIDTLFPNGVKI+AKF VGG + DSV+VPDILRMKNGVVP+Q I+  
 Sbjct: 298 IKYYVMVDFETFAAIDTLFPNGVKIDAKFATVGGVAVDSEVPDDLRMKNGVVPNQITE 357

40 Query: 332 VGIQVMDGRTLLNVARFRKDDGDGPORTRQQQVMRAIVSQIKDPRLEFTGSAAGKAYA 391  
 VG Q MDGRTLLNVARFRKDD+GDPORT RQQQVM A++SQIKDP +LFTGSAAGK YA  
 Sbjct: 358 VGBQMDGRTLLNVARFRKDDGDGPORTVRRQQQVMSAVMSQIKDPTKLTGSAAGIKIYA 417

45 Query: 392 LTSNLSYSFVLTGIPILSDAKNGIKMTIPREGDWDVDYQYGGQLTIDFKYKTL 451  
 LTS+N+S+ FV+ +G+ +L KNG++ +TIP GDWVD+YD YGGQ L IDF KY+K L  
 Sbjct: 418 LTSINVSFPFVVKNGVSVLGGKNGVHVHTIPENGWDVDEIMYGGQALYIDFKYQKTL 477

45 Query: 452 KSGLR 457  
 K+GLR  
 Sbjct: 478 AKLGLR 483

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 50 Example 1285

A DNA sequence (GBSx1362) was identified in *S. agalactiae* <SEQ ID 3951> which encodes the amino acid sequence <SEQ ID 3952>. This protein is predicted to be shikimate kinase (aroK). Analysis of this protein sequence reveals the following:

55 Possible site: 17  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

|                         |                                        |
|-------------------------|----------------------------------------|
| bacterial outside ---   | Certainty=0.3000 (Affirmative) < succ> |
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ>   |
| bacterial cytoplasm --- | Certainty=0.0000 (Not Clear) < succ>   |

60

-1421-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAAS5181 GB:X78413 shikimate kinase [Lactococcus lactis]  
Identities = 65/164 (39%), Positives = 98/164 (59%), Gaps = 8/164 (4%)

5 Query: 1 MPKVLGFMVGKTSVANCLENEVIDMDSLEIKHIGMSISRFFTEGEASPRALSESQFLN 60  
M +L+GFMG GK++VA L E D+D LIE+ I M I+ FF GEA PR +S++  
Sbjct: 1 MSIIILGFMGAGKSTVAKLLAREFTDLDKLIBEEIEMPIATFFELFGADPRKINIEVFE 60

10 Query: 61 ELLKKKNEGLVIAAGGGIVLLEENRRLLTLNRHNNIL-LTGSFEVLYHRIKKDBKRRPL 119  
++K ++IA+GGGI+ E + L L+R+ ++ LT F+ L+ RI D +N RP  
Sbjct: 61 LAVQK---DIIITATGGII--ENPKNLNLDRAASRVVFLTADFDTLWKRIEMDWQNVRP- 114

Query: 120 FLNHSKEEFYDIYQKRMILYSGLSDMIIDTYLTPQKIATVIGE 163  
L KE +++KRM YS ++D+ ID +P++IA I E  
15 Sbjct: 115 -LAQDKAAQLLFKKMKDYSLVADLTIDVTDKSPQIABQIRE 157

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3953> which encodes the amino acid sequence <SEQ ID 3954>. Analysis of this protein sequence reveals the following:

Possible site: 43  
20 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAAS5181 GB:X78413 shikimate kinase [Lactococcus lactis]  
Identities = 63/160 (39%), Positives = 97/160 (60%), Gaps = 5/160 (3%)

30 Query: 1 MTKVLGFMVGKTSVSKHLSMHCKDMDAIIFAKIGMSIAAFBQHGIEAFRTIESQVLK 60  
M+ +L+GFMG GK+TV+K L+ D+D +IE +I M IA FFE GE FR IE++V +  
Sbjct: 1 MSIIILGFMGAGKSTVAKLLAREFTDLDKLIBEEIEMPIATFFELFGADPRKINIEVFE 60

35 Query: 61 DLLFANDNSIIIVTGGVVVLQENRQLLRKNQHNNILLVASFETLYQRKHDKKSQRPLF 120  
L + II TGGG++ +N +L + + L A F+TL++R+ D ++ RP L  
Sbjct: 61 --LAVQKDIIITATGGGIENPKNLNLDRAASRVVFLTADFDTLWKRIEMDWQNVRP--L 115

40 Query: 121 KYSKEAFYEFYQQRMMVFYEGSLDILVIRVDRHTPEEVANII 160  
KEA +++RM Y ++DL I V ++FE++A I  
Sbjct: 116 AQDKAAQLLFKKMKDYSLVADLTIDVTDKSPQIABQI 155

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/161 (54%), Positives = 120/161 (73%), Gaps = 1/161 (0%)

45 Query: 1 MPKVLGFMVGKTSVANCLENEVIDMDSLEIKHIGMSISRFFTEGEASPRALSESQFLN 60  
N KVLGFMVGKTV+V+ L DMD++IE IGM+I FF + GE +FR +ESQ L  
Sbjct: 1 MTKVLGFMVGKTSVSKHLSMHCKDMDAIIFAKIGMSIAAFBQHGIEAFRTIESQVLK 60

50 Query: 61 ELLKKKNEGLVIAAGGGIVLLEENRRLLTLNRHNNILLTGSFEVLYHRIKKDBKRRPLF 120  
+LL N+ +I +GGG+V+L+RNR+LL N +NILL SFE LY R+K D+K++RPLF  
Sbjct: 61 DLLFA-NDNSIIIVTGGVVVLQENRQLLRKNQHNNILLVASFETLYQRKHDKKSQRPLF 119

55 Query: 121 LNHSSKEEFYDIYQKRMILYSGLSDMIIDTYLTPQKIATVI 161  
L +SKE FY+ YQ+RM+ Y GLSD++I D+ TP++A +I  
Sbjct: 120 LKYSKEAFYEFYQQRMMVFYEGSLDILVIRVDRHTPEEVANII 160

SEQ ID 3952 (GBS152) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 25 (lane 2; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 2; MW 45.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1286

A DNA sequence (GBSx1363) was identified in *S. agalactiae* <SEQ ID 3955> which encodes the amino acid sequence <SEQ ID 3956>. This protein is predicted to be 3-phosphoshikimate 1-carboxyvinyltransferase (aroA). Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.81 Transmembrane 241 - 257 (240 - 257)
INTEGRAL Likelihood = -0.06 Transmembrane 390 - 406 (390 - 406)

----- Final Results -----
 bacterial membrane --- Certainty=0.1723 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9673> which encodes amino acid sequence <SEQ ID 9674> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAD45819 GB:AF169483 5-enolpyruvylshikimate-3-phosphate synthase
[Streptococcus pneumoniae]
Identities = 288/426 (67%), Positives = 347/426 (80%)

Query: 5 MKLLTNANTLNGTIRVPGDKSISHRALIFGSIQGVTRIVDVLRGDEVLTIEAFKQWV 64
 NKL TN L G IRVPGDKSISHR+IIFGSI+G T++ D+LRGDEVLT++ F+ +GV
Sbjct: 1 MKLNTNIRHNLGIRVPGDKSISHRSIIFGSLAEGETKYDILRGDEVLTQVFFDLGV 60

Query: 65 LIEDDGBIITTYGKGFAGLTQPNLLDWMGNSGTSRLIAGVLAQGEFVTWVGDNLSLKR 124
 IED ++IT+ G G AGL P N L+MGNSGTS+RLI+GVLAG +FRV M GD+SLKR
Sbjct: 61 EIEDKDGVTIVQGVMAGLKAPQNALNMGNSGTSIKLISGVLAGDFRVMFGDLSLKR 120

Query: 125 PMDRIALPLSIQNGARISGVNDRPLPLKLGQTKKLPPIYHLPVASAQVKSALIFAALQT 184
 PMDR+ LPL IKG ISG T RDLPL+L+GK L+PI Y LP+ASAQVKSAL+FAALQ
Sbjct: 121 PMDRVTLPLKINGVSIQGTERTDLPLRLKGTGNLRPIHYELPIASAQVKSALMFAALCA 180

Query: 185 KGESLIVEKEQTRNHTEDMIQPGHLDIKDKIRLNGQQSLVQGDIRVPGDISSAAFVI 244
 KGES+I+EKE TRNHTEDM++QFGGHL + K+I + G Q L GQ + VPGDISSAAFV+
Sbjct: 181 KGESVIIKKEYTRNHTEDMLQFGGHLSDGKKITVQFGKLTQKVVVPGDISSAAFVL 240

Query: 245 VAGLIIPNSHIIENVGINETRTGILDVVSIMGGKIKLSVDNQVKSATLTVDSYHLQNT 304
 VAGLI FNS ++L+NVGINETRTGI+DV+ MGK++++ +D KEATL V+ S L+ T
Sbjct: 241 VAGLIAPNSRVLQNVGINETRTGIIIDVIRAMGGKLEITIEDPVAKSATLIVESSDLGTT 300

Query: 305 HISGAMIPRLIDELPIALLATQAQGTIVIAQAELEKVEETDRIQVVVESLQMGADITA 364
 I GA+IPLRIDELPIALLATQAQGTI TVI DA+ELKVEETDRIQVV ++L MGADIT
Sbjct: 301 EIGCALIPRLIDELPIALLATQAQGTIVIKDAEELKVEETDRIQVVADALNSMGADITP 360

Query: 365 TADGMIIRGNTPLHAASLDCHGDRIGMMTAAALIVKRGEDVLSGERAINTSYPNPLKH 424
 TADGMII+G + LH A ++ GDHRIGMM AIALLV +GEV+L RAINTSYP+ F +
Sbjct: 361 TADGMIIKGSALHGARVNIFGDHRIGMMTAAALIVADGEVSLDRARAINTSYPSFFD 420

Query: 425 LBLVN 430
 LE L++
Sbjct: 421 LESLIH 426

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3957> which encodes the amino acid sequence <SEQ ID 3958>. Analysis of this protein sequence reveals the following:

Possible site: 36

-1423-

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.18 Transmembrane 240 - 256 ( 239 - 256)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.1871 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- 10 >GP:AAD45819 GB:AF169483 5-enolpyruvylshikimate-3-phosphate synthase  
 [Streptococcus pneumoniae]  
 Identities = 278/426 (65%), Positives = 346/426 (80%)
- 15 Query: 4 MKLRINAGELQGTIQVPGDKSISHRAVILGAVAKGETRVKGLLKGEDVLSTIQAFNRLGV 63  
 MKL+TN L G I+VPGDKSISHR++I G++A+GET+V +L+GEDVLST+Q FR+LGV  
 Sbjct: 1 MKLKTINIRHLGIIRVPGDKSISHRSIIFGSLABGETKTVYDLRGEVDLSTMQVFRDLGV 60
- Query: 64 RIEEKDDQLVIEGQQFGGLNAPCQTIANMNSGTSMLRIAGLLAGQPPFSVMIGDESLSKR 123  
 IE+KD + ++G G GL AP LNMNSGTS+RLI+G+LAG F V+M GD+SLSKR  
 20 Sbjct: 61 EIEKDDGVITVQGVGAGLKAQRLIANMNSGTSIRLISGVLAGADFEVMPGDDSLSKR 120
- Query: 124 PMDRIVYPLKQMGVEISGETDRQFPFLQLQGNRNLPQITITLPISSAQVKSAILLAALQA 183  
 PMDR+ PLK+MGV ISG+T+R PFL+L+G +NL+PI Y LPI+SAQVKS++ AALQA  
 25 Sbjct: 121 PMDRVTLPLKMGVVISGQTERDLPFLRLKRTKRLRPIHYELPIASQAQVKSALMFAALQA 180
- Query: 184 KGTITQVVEKRTIRNHTERMIIQFGGRLVDGKRITLVGQQQLTAQETITVPGDISSAAPWL 243  
 KG + ++EKE TRNHTB+M+QQFGG L VDGG+IT+ GPQ+LT Q++ VPGDISSAAPWL  
 30 Sbjct: 181 KGESVIEKSTIRNHTERMIIQFGGRLVDGKKITVQGPQKLTQKRVVVGDISSAAPWL 240
- Query: 244 VAGLIIPGSELLKQNVGNPTRTGILEVVEKMGQAQIVTEMNKKEQVTSIRVVYSNMKGIT 303  
 VAGLI P S L+L+NVG+N TRTGIL++V+ MG ++ ++ + ++ V S++KGT  
 35 Sbjct: 241 VAGLIAPNSRLVNVGINETRTGILDIVIRAMGGKLEITLDPVAKSATLIVESSLKGIT 300
- Query: 304 IISGGILPRLIDELPIALLATQAQGTCTIKDAQELRVKETDRIQVUVTLNLSMGNKA 363  
 I G LIRLIDELPIALLATQAQS T IKDA+EL+VKETDRIQVV D LNSMGA+I  
 40 Sbjct: 301 EICGALPRLIDELPIALLATQAQGVTVIKDAELKVKETDRIQVADALNSMGADITP 360
- Query: 364 TADGMIIKGPVTLYGNNTSTYGDHRIGMNTAIALLVKQSQVHLDKBEAINTSYPTFFPD 423  
 TADGMIIKG + L+GA +T+GDHRIGMNTAIALLV G+V LE+ EAI TSYT+FF D  
 45 Sbjct: 361 TADGMIIKGSALHGARVNTFGDHRIGMNTAIALLVADGEVELDRAEAINTSYPTFFPD 420
- Query: 424 LERLCH 429  
 LE L H  
 Sbjct: 421 LESLH 426

An alignment of the GAS and GBS proteins is shown below.

Identities = 269/424 (63%), Positives = 331/424 (77%)

- 50 Query: 5 MKLLTNANTLKGITIRVPGDKSISHRAIFGSIQGVTRIVDVLRGEDVLSTIEAFKQNGV 64  
 MKL TNA L+GTL+VPGDKSISHRA+I G++++G TR+ +L+GEDVLSTI+AF+ +GV  
 Sbjct: 4 MKLRINAGELQGTIQVPGDKSISHRAVILGAVAKGETRVKGLLKGEDVLSTIQAFNRLGV 63
- Query: 65 LIEDDGEIITTYGKFGAGLTQPINLLMNSGTSMLRIAGLVLAQGFQEVTVMGDNLSEKR 124  
 IE+ + + I G+GF GL P L+MNSGTSMLRIAG+LAGQ F V M+GD SLSEKR  
 55 Sbjct: 64 RIEEKDDQLVIEGQQFGGLNAPCQTIANMNSGTSMLRIAGLLAGQPPFSVMIGDESLSKR 123
- Query: 125 PMDRIALPLSKMGARISGVTNRDLFPLKIQGTGKLLKPIFYHLPVASAQVKSALIFAALQT 184  
 PMDRI PL +MG ISG T+R PFL+LQG + L+PI Y LP++SAQVKS++ AALQA  
 60 Sbjct: 124 PMDRIVYPLKQMGVEISGETDRQFPFLQLQGNRNLPQITITLPISSAQVKSAILLAALQA 183
- Query: 185 KGESLIVEKRTIRNHTERMIRQGGHLIDIKKEIRINGGQSLVGQDITRVPGDISSAAPWI 244  
 KG + +VEKE TRNHTB+MI+QPGG L + K I L G Q L Q+I VPGDISSAAPW+  
 65 Sbjct: 184 KGTITQVVEKRTIRNHTERMIIQFGGRLVDGKRITLVGQQQLTAQETITVPGDISSAAPWL 243
- Query: 245 VAGLIIPNSHILENVGINETRTGILVDVSKMGGKIKLSSVDNQKSAATLVDYSHLQNT 304  
 VAGLIIP S ++L+NVG+N TRTGIL+V VMG +I ++ + + ++ V S++T

-1424-

Sbjct: 244 VAGLIIPGSEKLLKNNVGNPTRTGILEVVRKMGQAQIVYELMNKKEQVTSIRVVYSNNEGT 303  
 Query: 305 HISGAMIPLRIDELPIALLATQAQGTTVIADAQELVKVETDRICQVVSLEKMGADITA 364  
 ISG+IPLRIDELPIALLATQAQGT I DAQELVKETDRICQV + L MGA+I A  
 Sbjct: 304 IISGGLIPLRIDELPIALLATQAQGTTCIKDAQELRVKETDRICQVVTIDILNMGANIFA 363  
 Query: 365 TADGMIIRGNTPLHNASLDCGHDRIGMMIAIALLVKEGEVDLSGEEAINSTYPNPLEH 424  
 TADGMII+G T L+ A+ +GDHRI GMM IAIALLVK+G+V L ERAI TSYF F +  
 Sbjct: 364 TADGMIIRKGTPLVGANTSTYGDHRI GMMTAIALLVKQGVHLOKEAINTSYTPFFKD 423  
 Query: 425 LEGL 428  
 LE L  
 Sbjct: 424 LEEL 427

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1287

A DNA sequence (GBSx1364) was identified in *S.agalactiae* <SEQ ID 3959> which encodes the amino acid sequence <SEQ ID 3960>. Analysis of this protein sequence reveals the following:

20 Possible site: 38  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -1.12 Transmembrane 6 - 22 ( 6 - 22)  
 ----- Final Results -----  
 25 bacterial membrane --- Certainty=0.1447 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEP database.

30 >GP:AAF20148 GB:AF208390 actinin-like protein [Entamoeba  
 histolytica]  
 Identities = 62/236 (26%), Positives = 107/236 (45%), Gaps = 38/236 (16%)  
 Query: 144 NYHSTNSNPSGMLFYEQKIKWLSTH---IKYYLDYK--VTPIYQNNELIPKRIELK- 196  
 NN + N + + L N+++ N+ D+K V + + +I+ +  
 35 Sbjct: 116 NARQQRNVNAKEVVENNALDWNVSFGINVSFNSDWDGVALVKLTAVASAGQIKFEQ 175  
 Query: 197 YVGIDTKGKLLPIFIGNKSTQDQFGI-----STVLENTS PNATIDYLSGKAQN---- 245  
 + G+D T ++ K +CF I + E P++ Y+S +  
 40 Sbjct: 176 FSGLDNTQMWIDC---QLAYEQFKIPILMDVKDLVACERPDPKSIMITVSVYKERYQLL 232  
 Query: 246 TVLSAEQRKRLIAKHEERKRLAEK----KVEEKAARFTQKIL-BEQARLAARNAQ-RK 298  
 KE+++ IA+ F+E++ E+ + E+E A E Q+L L BEQ RIA E Q RK  
 45 Sbjct: 233 VEKQKEBQERLAREBQERKKEBQERLAREBQERLAREBQERLAREBQERLAREBQERK 292  
 Query: 299 QKEBQARLAARFTQKILKQKBTQVQSGYKRDYGRGWRHNPQVASKAELAAAGLQW 354  
 QKEBQ RIA E Q+++ QK+ Q + +P Q + + AA W.  
 Sbjct: 293 QKEBQERLAREBQERKQKEBQERLAREBQERLAREBQERLAREBQERLAREBQERK 338

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3961> which encodes the amino acid sequence <SEQ ID 3962>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have a cleavable N-term signal seq.  
 55 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1425-

The protein has homology with the following sequences in the databases:

```
>GP:CAA03161 GB:A49208 unnamed protein product [Streptococcus
pyogenes]
Identities = 54/222 (24%), Positives = 93/222 (41%), Gaps = 39/222 (17%)

5 Query: 44 HYQNTVSSKLLP--FTANYQLQLGELNLNRA-----TFSHIQLQRHETKQVTKINDY 96
 +YK +S++ P F + +LD L R T ++ ++ + K N +
Sbjct: 76 YYKTLGTSTQITPFLFPKAGDILYSKLDELGRITARTGLTYANVEGSGYVRSFGK-NQM 134

10 Query: 97 PVGWHN-----YQFFPGDG-SKSSWVNRGHLVGYQFGLNDEPNLVAMTAMLTGAY 149
 P GW Y+ + +G S NR HL+ G + + + A T
Sbjct: 135 PGMWTGNFNHVKYKIEWLNGLSYVGDFWNRSHLIADSLGG-----DALRVNAVNTGTRTQ 188

15 Query: 150 SGANDSNFHCMLYYNRLDGLWALHPDFMLDYKVTPISYSGNEVVEVRQIELQTVGIDSSGH 209
 + GM Y E R WL + D +L Y+V PIY+ +E++FR +
Sbjct: 189 NVGGRDQNGGWRITTSQRAQEWLEANKRDGYLYYEVAPITNADELIFRAV----- 236

Query: 210 LLTIRLNSNKNESIDENGVTTVILENSAPNIMLDVLANGTATPK 251
 + + S+ +I+E V++ N+A ++Y NGT T K
20 Sbjct: 237 --VSMQSSDNTINEK---VLVNTANGYITINANGTPTQK 272
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 117/245 (47%), Positives = 166/245 (67%), Gaps = 4/245 (1%)

25 Query: 2 KRKQPIKLGIAITLLTVISLYTPINLATNHTTENTIVTAQCY--KTKRNGTLFPKHKRQIVL 59
 K+K + + LL++ ++ A T N+ A + T + LPF QL L
Sbjct: 5 KQKASILTAVLLLSLSITTTVDRAKRVTPNVSHANTHYKNTVSSKLLPPTANTYQLQL 64

30 Query: 60 GELDDKGRATFAHIQLKVKDEPKKKVKRLKTTPVGWHNFKYFYNDGTQKAWLMSRGLI 119
 GELD+ RATF+HIQL+ + E K R K + PVGWHN+P Y DG++ +W+M+RG L+
Sbjct: 65 GELDNLRNATTFSHIQLQRHETKDVRTK-INYDPVGHWHYQFFYGDGSKSSWVNRGHLV 123

Query: 120 CHQFSGLNERRKNLVPMINWLTGYNSTNSSNPFESMLFYEKQLKTLWSTHKNYYLQYK 179
 +QF GLN+E +NLV MT WLNTG Y+ N SNPE ML+YE +L +WL+ H +++LDYKY
35 Sbjct: 124 GYQFGLNDEPNLVAMTAMLTGAYSGANDSNPEGMILYENRKLDSWLALHPDFMLDYKY 183

Query: 180 TPIYQNNELIPRKIELKYVIDKTRGLLPIFI-GNKSTQDQPGISTVTLENTSPNATIDY 238
 TPIY NE++PR+IEL+YUGID +G+LL I + NR + D+ G++IV LEN++PN +DY
Sbjct: 184 TPIYSGNEVVPRIELQYVGIDSSGELLITRLNSKNESIDENGVTTVILENSAPNINLDY 243

40 Query: 239 LSGKA 243
 L+G A
Sbjct: 244 LNGTA 248
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7263> which encodes amino acid sequence <SEQ ID 7264>. An alignment of the GAS and GBS sequences follows:

```
Score = 58.9 bits (140), Expect = 2e-11
Identities = 34/103 (33%), Positives = 55/103 (53%), Gaps = 1/103 (0%)

50 Query: 1 MPPKINLHAGILLYAMFMASITFILNLQVYLSCVTFALHKEYQACITDYKARLAEITVYQD- 59
 M K LKAGILL A+ +A++P LVLQ YL+++ A ++Y +Q+ KA L A++ Y+
Sbjct: 1 MTLKKLKAGILLQAIVLAAVFTPLVLQFYARILATERQVHSCIRASKAYILTAQLAYFTI 60

Query: 60 HRYKASNVVPPKGGQVICRERKKRWMLTVKLDQQRQYQFZYLK 102
 S + +P GG + + V LD+ Y+ ++ +
55 Sbjct: 61 EGDSSIGKCYPTGGYASYLQBGNYLQVKVTLDKGGWYNIHFPYR 103
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 1288

A DNA sequence (GBSx1365) was identified in *S.agalactiae* <SEQ ID 3963> which encodes the amino acid sequence <SEQ ID 3964>. This protein is predicted to be enolase (eno). Analysis of this protein sequence reveals the following:

```

5 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3025 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:BAA81815 GB:AB029313 enolase [Streptococcus intermedius]
 Identities = 396/435 (91%), Positives = 414/435 (95%), Gaps = 1/435 (0%)

 Query: 1 MSIIIDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVSLRDGDKSRYG 60
 MSIIIDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVSLRDGDKSRYG
 Sbjct: 1 MSIIIDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVSLRDGDKSRYG 60

20 Query: 61 GLGTQKAVDNVNNVIAEAIIIGYDVRDQCAIDRAMIALDGTPTNGKLGANAILGVSIIVAR 120
 GLGTQKAVDNVNNVIAEAIIIGYDVRDQCAIDRAMIALDGTPTNGKLGANAILGVSIIVAR
 Sbjct: 61 GLGTQKAVDNVNNVIAEAIIIGYDVRDQCAIDRAMIALDGTPTNGKLGANAILGVSIIVAR 120

25 Query: 121 AAADYLEFLYSLGGFTNKVLPFMQIIINGGSHSDAPIAQFQFMIPVGAFTFKEALR 180
 AAADYLEFLYSLGGFTNKVLPFMQIIINGGSHSDAPIAQFQFMIPVGAFTFKEALR
 Sbjct: 121 AAADYLEFLYSLGGFTNKVLPFMQIIINGGSHSDAPIAQFQFMIPVGAFTFKEALR 180

30 Query: 181 WGAIEVPHALKKILKRGLETVAGDEGGFAPKFBGTEDGVETILKATRAAGYRAGNGIMI 240
 WGAIEVPHALKKILK RGL TVAGDEGGFAPKFBGTEDGVETIL ATRAGY G++ ++
 Sbjct: 181 WGAIEVPHALKKILKSRGLATVAGDEGGFAPRFDGTEDGVETILAAIEAGYVFGIO-VFL 239

35 Query: 241 GFDCASEFYDIERKVYDYTSKFBEGGAVRTAARQIDYLEELVNKYPIITIEDGMQDENDW 300
 GFDCASEFYD ERKVYDY+KFBEGG AVRTA EQIDYLEELVNKYPIITIEDGMQDENDW
 Sbjct: 240 GFDCASEFYDKERKVYDYTKFBEGGAVRTADEQIDYLEELVNKYPIITIEDGMQDENDW 299

40 Query: 301 DGWKALTERLGGRVQLVGDDFFVTNTDYLARGIKKEAANSILIKVNQIGTLTTFEATEIM 360
 DGWK LTERLG +VQ VGDDFFVTNT YL +GI E ANSILIKVNQIGTLTTFEATEIM
 Sbjct: 300 DGWKALTERLGKIKVQVVGDDFFVTNTSLYLGKINEACANSILIKVNQIGTLTTFEATEIM 359

45 Query: 361 AKEAGYTAVSHRSGETEDSTTIADIAVATNAAGQIKTGSLSRTDRIAKYNQLLRIDQJLGE 420
 AKEAGYTAVSHRSGETEDSTTIADIAV NAQIKTGSLSRTDRIAKYNQLLRIDQJLGE
 Sbjct: 360 AKEAGYTAVSHRSGETEDSTTIADIAVANAGQIKTGSLSRTDRIAKYNQLLRIDQJLGE 419

50 Query: 421 VAQYHGKISFYNLKK 435
 VA+Y+G+KSFYNL K
 Sbjct: 420 VAQYHGKISFYNLKK 434

```

Proteins in the glycolysis/gluconeogenesis pathway have been experimentally detected on the surface of *Streptococci*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3965> which encodes the amino acid sequence <SEQ ID 3966>. Analysis of this protein sequence reveals the following:

```

55 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
60 bacterial cytoplasm --- Certainty=0.3025 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```



The protein has homology with the following sequences in the databases:

>GP:BAA81815 GB:AB029313 enolase [Streptococcus intermedius]  
Identities = 396/435 (91%), Positives = 415/435 (95%), Gaps = 1/435 (0%)

5 Query: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVLELDGDKSRYL 60  
MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVLELDGDKSRYL 60  
Sbjct: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVLELDGDKSRYL 60

10 Query: 61 GLGTQKAVDNNVNIAREAIIGYDVRDQQAIDRAMIALDGTPNKGKLGANAILGVSIAR 120  
GLGTQKAVDNNVNIAREAIIGYDVRDQQAIDRAMIALDGTPNKGKLGANAILGVSIAR 120  
Sbjct: 61 GLGTQKAVDNNVNIAREAIIGYDVRDQQAIDRAMIALDGTPNKGKLGANAILGVSIAR 120

15 Query: 121 AAADYLEVPLTYLGGFNTKVLPTPMNNIINGGSHSDAPIAQFPMIMFGAPTFKELR 180  
AAADYLEVPLTYLGGFNTKVLPTPMNNIINGGSHSDAPIAQFPMIMFGAPTFKELR 180  
Sbjct: 121 AAADYLEVPLTYLGGFNTKVLPTPMNNIINGGSHSDAPIAQFPMIMFGAPTFKELR 180

20 Query: 181 WGAEPHALKKILKERGLTAVDGGGFAKFBGCTDGVETILKAIEAGYEAENGIMI 240  
WGAEPHALKKILKERGLTAVDGGGFAKFBGCTDGVETILKAIEAGYEAENGIMI 240  
Sbjct: 181 WGAEPHALKKILKERGLTAVDGGGFAKFBGCTDGVETILKAIEAGYEAENGIMI 240

25 Query: 241 GFDCASSFYDKERKVDYTKFBGSGAAVRTSABQVDYLELVNKYPIITIEDGMENDW 300  
GFDCASSFYDKERKVDYTKFBGSGAAVRTSABQVDYLELVNKYPIITIEDGMENDW 300  
Sbjct: 241 GFDCASSFYDKERKVDYTKFBGSGAAVRTSABQVDYLELVNKYPIITIEDGMENDW 299

30 Query: 301 DGWKVLTSLRGLRVQLVGDFFVTINTYLARGIKENANSILIKVNGIGLTLTFPBAISM 360  
DGWKVLTSLRGLRVQLVGDFFVTINTYLARGIKENANSILIKVNGIGLTLTFPBAISM 360  
Sbjct: 301 DGWKVLTSLRGLRVQLVGDFFVTINTYLARGIKENANSILIKVNGIGLTLTFPBAISM 359

35 Query: 361 AKIAGTYAVVSHRSGTSDSTADIADVATNAGQIKTGSLSRTDRIAKNQILRIEDQLGE 420  
AKIAGTYAVVSHRSGTSDSTADIADVATNAGQIKTGSLSRTDRIAKNQILRIEDQLGE 420  
Sbjct: 361 AKIAGTYAVVSHRSGTSDSTADIADVATNAGQIKTGSLSRTDRIAKNQILRIEDQLGE 419

40 Query: 421 VAQYKGIKSPYNLKK 435  
VAQYKGIKSPYNLKK 435  
Sbjct: 421 VAQYKGIKSPYNLKK 434

An alignment of the GAS and GBS proteins is shown below.

Identities = 421/435 (96%), Positives = 427/435 (97%)

40 Query: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVLELDGDKSRYL 60  
MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVLELDGDKSRYL 60  
Sbjct: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVLELDGDKSRYL 60

45 Query: 61 GLGTQKAVDNNVNIAREAIIGYDVRDQQAIDRAMIALDGTPNKGKLGANAILGVSIAR 120  
GLGTQKAVDNNVNIAREAIIGYDVRDQQAIDRAMIALDGTPNKGKLGANAILGVSIAR 120  
Sbjct: 61 GLGTQKAVDNNVNIAREAIIGYDVRDQQAIDRAMIALDGTPNKGKLGANAILGVSIAR 120

50 Query: 121 AAADYLEVPLTYLGGFNTKVLPTPMNNIINGGSHSDAPIAQFPMIMFGAPTFKELR 180  
AAADYLEVPLTYLGGFNTKVLPTPMNNIINGGSHSDAPIAQFPMIMFGAPTFKELR 180  
Sbjct: 121 AAADYLEVPLTYLGGFNTKVLPTPMNNIINGGSHSDAPIAQFPMIMFGAPTFKELR 180

55 Query: 181 WGAEPHALKKILKERGLTAVDGGGFAKFBGCTDGVETILKAIEAGYEAENGIMI 240  
WGAEPHALKKILKERGLTAVDGGGFAKFBGCTDGVETILKAIEAGYEAENGIMI 240  
Sbjct: 181 WGAEPHALKKILKERGLTAVDGGGFAKFBGCTDGVETILKAIEAGYEAENGIMI 240

60 Query: 241 GFDCASSFYDKERKVDYTKFBGSGAAVRTSABQVDYLELVNKYPIITIEDGMENDW 300  
GFDCASSFYDKERKVDYTKFBGSGAAVRTSABQVDYLELVNKYPIITIEDGMENDW 300  
Sbjct: 241 GFDCASSFYDKERKVDYTKFBGSGAAVRTSABQVDYLELVNKYPIITIEDGMENDW 300

65 Query: 301 DGWKVLTSLRGLRVQLVGDFFVTINTYLARGIKENANSILIKVNGIGLTLTFPBAISM 360  
DGWKVLTSLRGLRVQLVGDFFVTINTYLARGIKENANSILIKVNGIGLTLTFPBAISM 360  
Sbjct: 301 DGWKVLTSLRGLRVQLVGDFFVTINTYLARGIKENANSILIKVNGIGLTLTFPBAISM 360

Query: 361 AKIAGTYAVVSHRSGTSDSTADIADVATNAGQIKTGSLSRTDRIAKNQILRIEDQLGE 420  
AKIAGTYAVVSHRSGTSDSTADIADVATNAGQIKTGSLSRTDRIAKNQILRIEDQLGE 420

-1428-

Seqjet: 361 AKEAGYIAVVSHRSGBTEDSTIADIAVATNAQGIKTPGSLSRDRIAKYNQLRIHQGLGE 420

Query: 421 VAQYKGIKSPYNLKK 435

VAQYKGIKSPYNLKK

5 Seqjet: 421 VAQYKGIKSPYNLKK 435

SEQ ID 3964 (GBS311) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 3; MW 51kDa).

GBS311-His was purified as shown in Figure 203, lane 10.

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1289

A DNA sequence (GBSx1366) was identified in *S.agalactiae* <SEQ ID 3967> which encodes the amino acid sequence <SEQ ID 3968>. Analysis of this protein sequence reveals the following:

15 Possible site: 60  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1998(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1290

A DNA sequence (GBSx1367) was identified in *S.agalactiae* <SEQ ID 3969> which encodes the amino acid sequence <SEQ ID 3970>. This protein is predicted to be di-/tripeptide transporter. Analysis of this protein sequence reveals the following:

30 Possible site: 54  
>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -14.33 | Transmembrane | 93 - 109 ( 87 - 122)   |
| INTEGRAL | Likelihood = -9.02  | Transmembrane | 117 - 133 ( 110 - 141) |
| INTEGRAL | Likelihood = -8.44  | Transmembrane | 333 - 349 ( 328 - 353) |
| INTEGRAL | Likelihood = -5.84  | Transmembrane | 19 - 35 ( 17 - 38)     |
| INTEGRAL | Likelihood = -3.08  | Transmembrane | 151 - 167 ( 151 - 167) |
| INTEGRAL | Likelihood = -2.55  | Transmembrane | 264 - 280 ( 264 - 281) |
| INTEGRAL | Likelihood = -2.28  | Transmembrane | 44 - 60 ( 44 - 60)     |
| INTEGRAL | Likelihood = -2.02  | Transmembrane | 238 - 254 ( 238 - 255) |

40 ----- Final Results -----

bacterial membrane --- Certainty=0.6731(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9395> which encodes amino acid sequence <SEQ ID 9396> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1429-

>GP:CB12175 GB:R99106 similar to di-tripeptide ABC transporter  
(membrane protein) [Bacillus subtilis]  
Identities = 175/359 (48%), Positives = 254/359 (70%), Gaps = 9/359 (2%)

5 Query: 1 MVMNLYGENDSRDAGPSIFVFCINLGAPISPIVVGVLQGVNPHLGPSLAATGMPGILL 60  
+VG+LY + D RRD+GPSIF GINLG ++P++VG LGQ+ N+HLGF AA+GM GL+  
Sbjct: 142 VVGDLYTKEDFRDSDGFSIFVFCINLGGLLAPLIVGTGLQKQYHNLGFGAAVGMGLGLI 201

10 Query: 61 QYTLGGKKYLTERRSLRPNDFSPERKSSLYKKVGLILIGIVIVILLIHAMMLTIEVLI 120  
+ L KK L +DLG +KS+ +G:I++ I +++ + +LTI+ ID  
Sbjct: 202 VVFLTRKRNGLAGSMVFNPLS--KKSAGTIGIGVITIAIAVISQ--TGVLTIRKRFID 257

15 Query: 121 IFSIIAIAIPPIIYPIKILSSKKISSVERSRVWAYIPLFIASILFWSIERQGSVVLALPAD 180  
+ SI+ I IP+IYFI + +SKR E+SR+ AY+PLFI +++FW+I+EQG+ +LA++AD  
Sbjct: 258 LVLSILGILIPVITYFIIMPTSKKADKTEKSRLAAYVPLFIQAVMFVAIQEGGATILANVAD 317

20 Query: 181 EQTKLNLNFFGHIHNPSSVFCSSNPFLIMLYVFPFAMLANQKSKQPSPPKKFAYGLFP 240  
E+ +L L F SS+FGS+NPLP++++ P FHLW K G +QPS+P KF+ G+  
Sbjct: 318 ERIRLSLGGF---ELQSSWFQSLNPLFVVIFAPIFANLMMKLGKRPSTPVKFSIGIIL 373

25 Query: 241 AGASFLAMMPLGLLPGFNAKVSFLWLTHSWAIVIVGEMLSIPVGLSNTSLKAPAFQAM 300  
AG SF+ M+ P + G A VSPFLWL +S+ +V++GR+ +SPVGLS T+KLAP AF AQ  
Sbjct: 374 AGLSFIIMVFPANQ-GKEALVSPLMLNLVSLFLVLVGLGLCLSPVGLSVTTKLAPAPASQT 432

30 Query: 301 MSIWFLSNAAQAQNAQIVKLYTPTDTOTLYYGVVGITVVPFILLFYVPRIEKLSGV 359  
MS+WFL+NAQAQINAQ+ L+ +T+Y+G +G I++V G ILL P I++ M GV  
Sbjct: 433 MSWFLNAAQAQNAQVAGLFDKIPETMYFCTGLISIVLGILLLSPVIKRAMKGV 491

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1291

A DNA sequence (GBSx1369) was identified in *S.agalactiae* <SEQ ID 3971> which encodes the amino acid sequence <SEQ ID 3972>. Analysis of this protein sequence reveals the following:

35 Possible site: 37  
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

40 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1292

A DNA sequence (GBSx1370) was identified in *S.agalactiae* <SEQ ID 3973> which encodes the amino acid sequence <SEQ ID 3974>. Analysis of this protein sequence reveals the following:

50 Possible site: 30  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2485 (Affirmative) < succ>

-1430-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF61315 GB:U96166 unknown [Streptococcus cristatus]  
 Identities = 181/442 (40%), Positives = 270/442 (60%), Gaps = 2/442 (0%)

Query: 1 MINLEDSYTGSSWDLHFLKSGYINPTIALNDDGLFDQVTSPLYLYTTFKNGAGRP 60  
 10 MI LFD Y Q+S+DL SL +G P + + DDG+L DV SPY Y+TG T GRP+  
 Sbjct: 1 MICLEFDYQASFDLLRSLKATGLDCPVVVQDDGYLSFDVSPYKSYFTGLDTPPEGRPT 60

Query: 61 YYNELRVEDTWELIGFSSGADIVLGVKGRITIANPNHKKLKEVDFDQGRVILHGR 120  
 Y+N + P WEI + +I+D+G K+ I Y P H+R ++ V+W D +G+V D  
 15 Sbjct: 61 YFNLFKPHLWEIRSSNVNGHLLMGKKRANIPYRQPTHERRRVAVEMLOTGGQVRAAD 120

Query: 121 FNNKGFCTAQTTFINADGQAIQTSYNNKDRJEVISENNHMGDIYLLNDRNPKVYFKSKVLFV 180  
 +N+ G FAQ Y+ + T Y+++ VI KWH+TSD IL + +FKSK BFF  
 15 Sbjct: 121 YNNKGRLEFAQITYDQTRPFTHTYFDQSNVVMENHLLTGLDILTLBGRKHIFKSKQBFV 180

Query: 181 INYLQEAQFNLDRIFFNLSLTPEPLVSFYL--NRLESKDVLFQWJEFVDDIPGNMRLLIAN 238  
 + YLQ ++ DRI YNLSLTPEPLV++ L ++DVLFWQJE++ +PGNW++ +  
 20 Sbjct: 181 VFYLQIRGYDTDIRIYNLSLTPEPLVAYALRPKNGRAEDVLFQWJEIIGALPGNPKVAMQI 240

Query: 239 DSRNTKIVQSYEAYANMRLTDEEQKQVSPFLGFWYFKETKLNQALILTNSDQIEA 298  
 P N +I +Q + Y L T EE+ +G++Y + + +ALILTNSDQ+E  
 25 Sbjct: 241 PERNIRIAVQDRQVYKIQSLATPEEKVYFENIGYTYDQRLNNNEALILTNSDQIEQ 300

Query: 299 LESLIVSLPLNLFNIGALTEMSDLANFGKYNVNVLYNITNQIQLYSNICAFYLDINH 358  
 +E L+T LPS+ P+IGA+TEMS LM +Y NV LYNI ++ L C YLDIN  
 30 Sbjct: 301 IEQLLTQLRNVPHIGAITEMSGHLMGLNRYENVSLYNIRPAKVAELFERDQVLYDINI 360

Query: 359 HNEILSRVASPEHQQLIFAPEETSHQIRFVSQNIHFKKDIPTTSHLQPLIGNKCNIE 418  
 +EIL+A R+APE+ LI +F T H RF++ +I+ ++ + +Q + + +E  
 35 Sbjct: 361 SDEILNACTAPENNMLILSPFTWCHSRFPADDDHIAPEENVSGMVKIQSALAHSSME 420

Query: 419 KALKKQLEDCHVSSSTQYQSVI 440  
 AL +Q + + +S QY++I  
 Sbjct: 421 ANITRQKQANQASLECYKAI 442

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1293

A DNA sequence (GBSx1371) was identified in *S.agalactiae* <SEQ ID 3975> which encodes the amino acid sequence <SEQ ID 3976>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.06 Transmembrane 405 - 421 ( 404 - 422)

50 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1022 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA094320 GB:AB033763 hypothetical protein [Staphylococcus aureus]  
 Identities = 66/195 (33%), Positives = 99/195 (49%), Gaps = 9/195 (4%)

-1431-

Query: 259 NYDYQFTNANRFDDFTTSTKQIKELLSQQKQFTNNHRIITIPVGSID---NLKMF 314  
 N Y + F N NR+ I ST +Q + N+ TIPVG ID + NLK  
 Sbjct: 15 NTKYHVFNNLNRYSGIIVSTKQQ---QLDISARINNEIPVHTIPVGYIDEHFTNLKNN 70

5 Query: 315 DNRRPYSILTSASRLASEKHVDWLVRVIRIREILPEVTFDIYSGSGSEKHINLINANA 374  
 + I++ +R +EK ++ + V++ + P + +YG G EEEK +I N  
 Sbjct: 71 HSTNNKLIISVARYSPERQNLHQIQLVSKLKEPFRILHLIYGFCKEESBYKQLITEYNL 130

10 Query: 375 TEYIRLMG-HQSLGNVQNYSLYLITASKSEGFGLITLRAIGAGLPLIGFDVRYGNTFFIK 433  
 + L G +NLG Q+ + L S EGF L LLE I G+P +G++ +YG I  
 Sbjct: 131 ENNVFLGFRNRLGAEIQAYMSLITENMEGFNLGLLITTEGIPFVYNSKYGPGSELIL 190

Query: 434 DGENGYLIFRFMD 448  
 + ENGYLI + D D+  
 15 Sbjct: 191 NNGENGYLINKMKDE 205

SEQ ID 3976 (GBS426) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 4; MW 58.8kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 3; MW 84kDa).

20 GBS426-GST was purified as shown in Figure 220, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1294

A DNA sequence (GBSx1372) was identified in *S. agalactiae* <SEQ ID 3977> which encodes the amino acid sequence <SEQ ID 3978>. This protein is predicted to be preprotein translocase *seca* subunit (*seca*). Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.69 Transmembrane 75 - 91 ( 75 - 91)

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1277 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA04957 GB:U56901 involved in protein export [Bacillus subtilis]  
 Identities = 336/794 (42%), Positives = 506/794 (63%), Gaps = 29/794 (3%)

40 Query: 5 NSLPSLDKKRLKQLQRTINTINSILKGMATLSNRELQAKTTFPRKRLVNGETLDDICAE 64  
 N +F K+ L + ++ N I++++G LS++ L+ KT EF+RL G T DD+ EA  
 Sbjct: 6 NKMFDPTRKRLNRYEKIANDIDAARDYENLSDDALKHKITTEPRKRLKGGATTDLLVEA 65

45 Query: 65 FAVVRADRRVLGLFPYDVQVIGGLVLHQSGNTAEKMTGEGKTLTATMPYLALAEKSGM 124  
 FAVVREA RV G+FP+ VQ++GG+ LH GN ARMTGEGKTLT+T+P+YLNALG EKG  
 Sbjct: 66 FAVVRASRRVYGMFPFKVQLMGVALIDGNIAEMTKGEGKTLTSTLPVYLNALTKGVH 125

Query: 125 LLTNSYLAIRDAREMGKVYRPLGLSVGVGVSINEREDRDAATKRAVYSDDIVYSTSAL 184  
 ++T N YLA RDAB+MGK++ FGL+VG+ ++ ++ KR Y++DI YST++ L  
 50 Sbjct: 126 VVTVNEYLASRDABMGKIFEPGLITVGLINLNSMKSDE----KRAYAADITYSTNDEL 180

Query: 185 GFDYLLINLASSKQKQKMPKLYHVAIVDEADVLDDMAQTPLVSGSPRVQSNLYKIADEL 244  
 GFDYL DN+ K Q LH+AA++DE D++L+D A+TPL+ISG + LY A+  
 Sbjct: 181 GFDYLRDNNVLYKQMVQRPILHFAVIDEVDGILIDEARTPLILISGQAASKTLVQANAF 240

55 Query: 245 ILSFEQVDVYFDKREQVEWIKNQGVREARYRIFPHFYQKSNRELVRHLNLSKAKHLF 304  
 + + + + DY +D + + V + +G+ +AE+ F I + + + L H+N +LKH  
 Sbjct: 241 VRTLKAEKDYTYDIKTAKVQLTRGNTKARKAPGIDNLFDVKRVALNHHINQALKAHVM 300

-1432-

Query: 305 ERGKDYVVDGGEIKLDAITNGVLEGTQKGGVHQATBQKEHLAVTPESRAMASITYQN 364  
 ++ DVVV+DG++ ++D+ GR+++G+ G+HQATE KE L+ ES +A+IT+QN  
 Sbjet: 301 QKIVDYVYVREDGQVIVDSFTGRIMKGRKYSEGLHQATRAKEGLEITQNESMTLATITPQNY 360

Query: 365 FRMF+TKLAGMTGTGKTAKEKFBFVYDMSVVRIPINSFVRRIDYEDKIYTYLPEKIHATIE 424  
 FRM+ KLAGMTGT KT E+EF +Y+M+VV IPIN FV R D ED IY T+ K A E  
 Sbjet: 361 FRMYEKLAGMTGTAKTESEBFFNITYMQVVTIPNRFVVRDDRDPDIYKTMGKGKFAVAE 420

Query: 425 FVKKVHDTGQPTLLVAGSVMSSELFSLELLSGIPHSILNAQSAVKEAQMLAEAGQKGAV 484  
 V Q + TQGP+L+ +V SEL S+LL GIPH +LNA++ +HAQ+I EAQQKQAV  
 Sbjet: 421 DVAKRYMTGQPVLVGTVAVETSELISKLKNGKIPHQVLNAKNHEREAQILIEAGQKGAV 480

Query: 485 TVATNMAGRGTIDIKLGKGVSELGLAVIGTERMKSRQMDQLAGRSRQGDIGFSQFFVS 544  
 T+ATNMAGRGTIDIKLG+GV ELGGLAV+GTER +S+R+D QLGRSRSRQGD G+QF++S  
 Sbjet: 481 TIATNMAGRGTIDIKLGBGVSELGLAVVGTERRERSRIDNQLGRSRSRQGDPGITQFTLS 540

Query: 545 FEDDLMIRSGPKNAQDYFRKNRDKVNPFKALGQRRPQKLFQCTQEARSDGKESARSQT 604  
 ED+LM G + D+ + + + + +Q+ +G +R Q  
 Sbjet: 541 NEDELNRFRGAERTMAML---DRFGDDSTFIQSDMVSRVSSQKRVBGNFNFSRQL 596

Query: 605 IEPDSSVQLRETYVYRERNALINGESGHFSRQIITDVISSFT----AYLGEVSEKEEL 659  
 +++D ++ QRE +Y++R +I+ E + R+I++ +I S + AY E EE  
 Sbjet: 597 LQYDVLRAQQRREVIYKQRFVIDSE---NLREIVEMIKSSLEKRAIAAITYPRELPEE- 651

Query: 660 IFVNNRPI-FDNMSYNLQGISKEMSL--BEIKNYLFKIADSIILREKHNLLGDSFG----- 711  
 ++++ + N+Y +G ++ + +E L I D I+ K+N + FG  
 Sbjet: 652 -WKLGSLVDLINTYTLDEGALKSDIFQKEPDENLELIDRII-TKYNEKEQPGKEQNR 709

Query: 712 DFERTAALKADEAWIEVEDVYLQQLATVATARQTAQRNPFVEZHKEAYKSYNIMKKEIRE 771  
 +FE+ L+A+D W++ +D + QLR E AQ NP+ EY E + + M + I +  
 Sbjet: 710 EFERIVTLRAVDSKMMHDIDAMDQLRQIHLRAYAQVNLREYQMGGFAMFHSIMEST 769

Query: 772 QTFRNLLSEVSPN 785  
 + + ++ +E+ N  
 Sbjet: 770 EVAKFVMKBIENN 783

There is also homology to SEQ ID 3620.

SEQ ID 3978 (GBS425) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 3; MW 91kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 2; MW 116kDa).

GBS425-GST was purified as shown in Figure 220, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1295

A DNA sequence (GBSx1373) was identified in *S. agalactiae* <SEQ ID 3979> which encodes the amino acid sequence <SEQ ID 3980>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3827 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1433-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1296

- 5 A DNA sequence (GBSx1374) was identified in *S.agalactiae* <SEQ ID 3981> which encodes the amino acid sequence <SEQ ID 3982>. Analysis of this protein sequence reveals the following:

Possible site: 17  
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2683 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 10001> which encodes amino acid sequence <SEQ ID 10002> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1297

A DNA sequence (GBSx1375) was identified in *S.agalactiae* <SEQ ID 3983> which encodes the amino acid sequence <SEQ ID 3984>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.5410 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1298

A DNA sequence (GBSx1376) was identified in *S.agalactiae* <SEQ ID 3985> which encodes the amino acid sequence <SEQ ID 3986>. This protein is predicted to be preprotein translocase secY subunit. Analysis of this protein sequence reveals the following:

40 Possible site: 59  
>>> Seems to have an uncleavable N-term signal seq  
INTEGRAL Likelihood = -9.92 Transmembrane 287 - 303 ( 278 - 309)  
INTEGRAL Likelihood = -9.08 Transmembrane 191 - 207 ( 186 - 210)  
INTEGRAL Likelihood = -8.44 Transmembrane 104 - 120 ( 101 - 123)  
45 INTEGRAL Likelihood = -8.23 Transmembrane 11 - 27 ( 9 - 41)

-1434-

```

INTEGRAL Likelihood = -3.93 Transmembrane 133 - 149 (129 - 150)
INTEGRAL Likelihood = -3.19 Transmembrane 347 - 363 (344 - 364)
INTEGRAL Likelihood = -2.97 Transmembrane 158 - 174 (155 - 174)
INTEGRAL Likelihood = -1.54 Transmembrane 246 - 262 (245 - 262)
INTEGRAL Likelihood = -0.90 Transmembrane 372 - 388 (372 - 388)
INTEGRAL Likelihood = -0.85 Transmembrane 64 - 80 (64 - 81)

----- Final Results -----
bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF30659 GB:AE002122 preprotein translocase [Ureaplasma urealyticum]
Identities = 105/422 (24%), Positives = 213/422 (49%), Gaps = 49/422 (11%)

Query: 2 KLLYIFERNIILRKILITFSLIIIFLLGRYVPIQVLIISAYKQDNNFATLYSTVTGNNL 61
+LL IF+ +L +++T S++I+F +G +P+P + ++ G +F ++ ++ G3 L
Sbjct: 13 QLLMIFKNKVLVALIVLISLILFRIGSVIPMPYIKLNGNFGQSGFSSIINLGG3GL 72

Query: 62 SQVGFSLGIGPMITMILLRFT-----IGKYSSGVSVQKQVQFRNVMLVIAII 112
SQ +P+GIGP +T I++L+ + + K +K++ + ++ L ++
Sbjct: 73 SQFSLFAIGIGPYITAIQIMQLSSELVPLLAKLSKSGGRKKIEVITR-IITLPLAVN 131

Query: 113 QGLAITISFQYHNGFSL-----TKLLLATMI--LVTGAYIISWIGNLNAEYGP3- 159
Q + I NGF + L T I +V G YI ++ +L ++ G G
Sbjct: 132 QAVIINLMTRANQFISIVSNAPFAGISPLFYTYIFLMVGQTYISLPLADLISKKGVN 191

Query: 160 GMTILVVGMINGQFNNIPLIFELF-----QDGYQLAIIPLMLLVAMYIMITPERSE 213
G+T+L++ G++ FN: IF + + IL+L+ ++ + ++ S
Sbjct: 192 GITLLILTGIVASLNFHETAFISNGLSLSSTSSKVSQIIGFILYILFYIMILGIVFVNST 251

Query: 214 YRIPVMTS-----IHNRLVEDAMPYIKVNASGNAFNVYITLMPFQYIIILLRSIFPT 268
+IPV +T H +L ++PIK+ +G M ++ ++L P + L
Sbjct: 252 RKIPVQQTQALILDHEKL-----PFLPIKMTAGVMPVIFASVLAIRPQVAEFLDK--Q 305

Query: 269 NPDITSYNDYFSLSSIGQVVIYIMILVLVLSVAFIVMDPTKISEMRESGDFINPYR3 328
+ ++YF + S G+ IY++L+L+ + F++V ++P K++E +++G FIP + G
Sbjct: 306 SMGYVIHNYFIYDVSWTGLATYVVLILLFTFFFSYVQLMPFKGARDIKGRFIRPGV3 365

Query: 329 KETQSVLSKICYLPGTFSGFFMAFLG3VFLFALGNDLDR-----TVSSMTGIRFM 379
+T+ +++K+ Y +AFL +P L AL + T+ T I +M
Sbjct: 366 MDETHGHTKVYRVNMIGAPILAPLACLPLHVALVAKTINHGIPIVQPSITPGT3IIM 425

Query: 380 IT 381
+T
Sbjct: 426 VT 427

```

There is also homology to SEQ ID 3988.

A related GBS gene <SEQ ID 8783> and protein <SEQ ID 8784> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: 6.32
GVH: Signal Score (-7.5): -4.07
Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 10 value: -9.92 threshold: 0.0
INTEGRAL Likelihood = -9.92 Transmembrane 287 - 303 (278 - 309)
INTEGRAL Likelihood = -9.08 Transmembrane 191 - 207 (186 - 210)
INTEGRAL Likelihood = -8.44 Transmembrane 104 - 120 (101 - 123)
INTEGRAL Likelihood = -8.23 Transmembrane 11 - 27 (9 - 41)
INTEGRAL Likelihood = -3.93 Transmembrane 133 - 149 (129 - 150)
INTEGRAL Likelihood = -3.19 Transmembrane 347 - 363 (344 - 364)
INTEGRAL Likelihood = -2.97 Transmembrane 158 - 174 (155 - 174)

```



-1435-

```

INTEGRAL likelihood = -1.54 Transmembrane 246 - 262 (245 - 262)
INTEGRAL likelihood = -0.90 Transmembrane 372 - 388 (372 - 388)
INTEGRAL likelihood = -0.85 Transmembrane 64 - 80 (64 - 81)
PERIPHERAL likelihood = 8.65 28
modified ALOM score: 2.48

```

\*\*\* Reasoning Step: 3

----- Final Results -----

```

10 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

15 RQPD2350(316 - 1500 of 1277)
 GRAB[6621]6420(8 - 426 of 431) preprotein translocase secY subunit [Bacillus sp.]
 SP[38375]SECY_BACDH PREPROTEIN TRANSLOCASE SECY SUBUNIT. GP[484251]dbj|BAA01191.1||D10360
 secretion protein Y [Bacillus sp.] PIR[B44859]B44859 preprotein translocase secY - Bacillus
20 sp.
 %Match = 12.1
 %Identity = 26.8 %Similarity = 55.4
 Matches = 109 Mismatches = 165 Conservative Sub.s = 116

```

57 87 117 147 177 207 237 267  
25 EVQNVVDRCTIGKTTYGIRPARKDNCYISFERIMDDFEYLCDTIKNR\*SRRVMT\*ILKSIPLILKLT\*LT\*SYLG\*

297            327            357            387                         441            471            501  
REQIDREIPLKLLYIFENIIIRKLILTFSLIIIFLLGRYPVIGV--LISAYKGQDNNAFLYSTVTGGNLSGVF  
| | : | : | : | : | : | : | : | : | : | : | : | : | : |  
MFTIINIFPVCDLRKRVFIPTLMVLIVRIGSPFIVPOTNEVLDVFDQMAFGEL-NTFGGGLNGNSPI

30            10            20            30            40            50            60            70

531                    582                    594                    624                    654                    681                    699

SLGIGPMMTMLLRLLF---TIGKYSSGVSC-----KVQCFNMVVLIAIIQGLATISIFQ-YHNGF-----SLTKLL

||| | : ||:: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

AMGINPIFTASIVMLQLQMVDVPKFAEKAGEAGRRGLAOPTRYTYTVLVLGIQALMGSGVFNPFPFGLLPNPSVSFL

80                    90                    100                    110                    120                    130                    140                    150

[illegible]

```

45 966 1023 1053 1083 1113 1143
ITPERSEYR-IPUM---RTISHNRLLVDA-YMPIKVNAGSGMAFMYVYTLRLMPQYIILLRSIFPTNPDITSYNDYFSL
: : : | | | : : : : : | | | : : | | | : : | | : : : | |
VIFVQQLAKLIPQVAKRLVGNRPMGGGSLPLKLNAGGVIPIFALSILIPPTVAGLFGSDHVAIAWETDYF---

```

[illegible]

60

1413      1440      1470      1500      1530      1560      1590      1620  
LGNDLRTVSSMTGI-FMMITGNSFMILDEQVIRIKRKYTSVFENEEN\*CFILFILGIMKIVLGLMIITGGISSRLMSV  
| | : : : : : : : : : : : : : : : :  
IKFADLPQAIGISTGLLVGVVALTKMQLEQLKRSGKPIK  
400            410            420            430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1436-

**Example 1299**

A DNA sequence (GBSx1377) was identified in *S.agalactiae* <SEQ ID 3989> which encodes the amino acid sequence <SEQ ID 3990>. Analysis of this protein sequence reveals the following:

Possible site: 24  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3002 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF61315 GB:U96166 unknown [Streptococcus cristatus]  
 Identities = 30/78 (38%), Positives = 41/78 (52%)

Query: 276 ALTVTLTDDIMELEHLLQRCNPTDFHIAAPVYCSDRKQWGYPNYYLHEAITEQEVL 335  
           AL +T +D + ++E LL + PN FHI A S L L YPN L+ I + L  
 Sbjct: 289 ALILTNSDQLEQIQLLTQLPNVHFHGAITEMSGHLAQLNRYPNVSLYFNIRPAKVAEL 348

Query: 336 LLNSDIYLDINHGEEVWN 353  
           D+YLDIN +E+ N  
 Sbjct: 349 FERCDLYLDINISDEILN 366

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1300**

A DNA sequence (GBSx1378) was identified in *S.agalactiae* <SEQ ID 3991> which encodes the amino acid sequence <SEQ ID 3992>. This protein is predicted to be eps7. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> May be a lipoprotein

----- Final Results -----  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC07458 GB:AX009404 product = eps7 [Streptococcus thermophilus]  
 Identities = 87/232 (37%), Positives = 133/232 (56%), Gaps = 22/232 (9%)

Query: 10 VSVIIPVYNAAPYLEGCVNTFILOQTYQVFERILLIDDGSTDVTSASICDQLSLRDNIRVFH 69  
           S++IPVYN Y++ C+++IL QT+ EI+L+DDGSTD S ICD S D RI+V H  
 Sbjct: 3 ISIVIPVYVQDYKKCLDILSQTFSLEIILVDDGSTDLSGRICDYISENDKRIKVIH 62

Query: 70 IENGASKARNFGLARISPSQFVTVFVSDDWKENYLEVLLAQOEKNADIVISNYIY 129  
           NGG S+AFN G+ + S+++TF+DSD+V +Y+E L + +NADI I+++  
 Sbjct: 63 TANGCQGEARNVGLKNAT--SEWITFIDSDYVSSDYIEVLYNLQVHNADISIASF--- 117

Query: 130 RETEDIFGYITDKDFV-----IKEISQATIDROVHMLNSSVFVINGKLYRRELF 183  
           YIT K + + + A+TAI R + IN + +WKG+YR E F+  
 Sbjct: 118 -----TYITPKLIKHGNGEVALMDAKTAIRRL-----LNEGFMGVWGMKRYETFN 166

Query: 184 TITFPIDKVFEDELVSVLLFIKSKKFIIVNGSYGYRIRFNSIMTSAFSSKR 235  
           F K+PED L++ +F ++ + Y Y R NS + F+ K+

-1437-

Subject: 167 KYKFGVSGKLPEDSLITYQFSEASTIVFGADLYFYVNRKKNSTVNGTFNKK 218

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 5 Example 1301

A DNA sequence (GBSx1379) was identified in *S.agalactiae* <SEQ ID 3993> which encodes the amino acid sequence <SEQ ID 3994>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1569(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 20 Example 1302

A DNA sequence (GBSx1380) was identified in *S.agalactiae* <SEQ ID 3995> which encodes the amino acid sequence <SEQ ID 3996>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1662(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 35 Example 1303

A DNA sequence (GBSx1381) was identified in *S.agalactiae* <SEQ ID 3997> which encodes the amino acid sequence <SEQ ID 3998>. This protein is predicted to be a glycosyl transferase (gspA). Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2606(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-1438-

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus ducreyi]  
 Identities = 62/177 (35%), Positives = 105/177 (59%), Gaps = 8/177 (4%)

5 Query: 3 YARYYIPQLIDAEKVLYLDDITLVVDNLOKLFETELGQYPTAAILD--GDGIY----FN 55  
 + RY+I I+ +KV+YLD D +V +L +L+ + +Y +AA+ D + IY FN  
 Sbjct: 69 FFRFYISDFTEODKVIYLDADIVVNGSLTELYCTDISNYFLAAVKDIISEKIYVNNHFN 148

10 Query: 56 SGVMLNSLYWMRYRVTEKLLKITEKELONGIFGQQGVNLFFDNNMLKLEDKYNAQVGN 115  
 +G++LIN+ W + +T+ L ++E+ +++ DQ +LIN+P+ + WRL YN +G  
 Sbjct: 149 AGMLLNNKRWREHNTITFCLSLSKRYNSLPDAQSILNLIFPKRWGLKLNRYNYLIGT 208

Query: 116 DLGAFYENWQGYFDRNFES--PTIIHYCTHDKPNTFSSSRPFRKTWQYBQLDWEVFP 171  
 D F Y + E+ P IHHY T KFW ++RFR +W Y +L+W +++  
 15 Sbjct: 209 DYLPFKYGRTHYLEDLGETIPLIHHYNTKRWKLNITFRFRNIYWFYELANWQDIY 265

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 20 Example 1304

A DNA sequence (GBSx1384) was identified in *S.agalactiae* <SEQ ID 3999> which encodes the amino acid sequence <SEQ ID 4000>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

25 Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1157 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus ducreyi]  
 Identities = 103/259 (39%), Positives = 156/259 (59%), Gaps = 3/259 (1%)

35 Query: 7 IALAADPGYQGVKTIKISICPHNQFIDFYILNDDFFVFNQWMEYHLSKMDCTISNYKI 66  
 I LAA+ Y E+ T IKSI HN+ I FY+LN D+P EWF ++ L K+ I + K+  
 Sbjct: 10 IYLAANQSYSEYILITIKSIYILNKHIRFYLLNPDYPTWFDILNKLRLKLNSEIDIKY 69

40 Query: 67 FNEELKEPK-FQKMPFYPTFYFRYPVEVIEHDKVLYLDCMLITSDLTISFTLDISKGV 125  
 N+ IK+FK + T+FRYFI + I+DRV+YLD D+++ LT ++ DIS Y +  
 Sbjct: 70 TNUITRNKFTYSHISSDTTFPRYISDFIEQKVIYLDADIVVNGSLTELYQIDISNYPL 129

45 Query: 126 AAVRDDLEEYDGKEYDFNSGLLLINNITFWRBQGISRLLYTYRNGQALQYHCDQVIND 185  
 AAV+D + E+ FM+G+LLINN WRE I+Q L + + +L DQ +LN  
 Sbjct: 130 AAVKDIISEKIYVNNHIFMAGMLLNNKRWREHNTITFCLSLSKRYNSLPDAQSILNL 189

50 Query: 186 VICJNWLELDETYNHYTGADMLYNLQSQSRQWRRKDLKPKVHYH-TATKPKYLETSVR 244  
 + D WL+L+ YNY G D L+ + ++ + +P +HHY T KFW + + R  
 Sbjct: 190 IFDKRWGLKLNRYNYLIGTDYLPFKYGRTHYLEDLGETIPLIHHYNTKRWKLNIT-FNTR 248

Query: 245 WRDIWWEYNRLSWRDIFTR 263  
 +R+I+W Y L W+DI+ +  
 55 Sbjct: 249 FRNIYWFYELANWQDIYAK 267

No corresponding DNA sequence was identified in *S.pyogenes*.

-1439-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1305**

A DNA sequence (GBSx1385) was identified in *S. agalactiae* <SEQ ID 4001> which encodes the amino acid sequence <SEQ ID 4002>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 28  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2679 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus  
ducreyi]  
Identities = 94/263 (35%), Positives = 158/263 (59%), Gaps = 4/263 (1%)

Query: 2 KKTIVLGADFOYRDQVMTIKSI VSHNQHLTIYIINTDFVEMFNLNHSLEQFDCRVN 61  
K IVL A+ Y + ++TTIKSI HN+H+ Y+N D+P EMF+ILN+ L + + + +  
Sbjct: 7 KGNIVLAANQSYSEYILATTIKSIYLNKHIRFYLNDRDYPTEWFDILNNKLRKLNSEIID 66

Query: 62 IPISSDVFGPIPTLSHISV-AGFFRWFIPIHLKEEIVLYLDSQVIRGSLDPLFDINLES 120  
I +++D + T SHIS FFR+FI +E++ V+YLD+D++V GSL L+ + + +  
Sbjct: 67 IKVINTDIKNFKTYSHISSDTFFRYFISDFIEQDKVIYLDADIVNGSLTELYQDIDSN 126

Query: 121 NLLGAVADHPSLTLYYGDTPAPVSFNGVMLINNSLWKEEIVNLSMRIADKG-SAVGVGDQ 179  
L AV D S Y + FN+G++LINN W++ I + ++K +++ DQ  
Sbjct: 127 YFLAAVKDIISEKIYVNNH--IFNAGMLLINNKKVREHNITQFCLSSEKYNLSPLDAQ 184

Query: 180 EYLNILITQNRWIDIGQYQVQIGQOVNINAYGRPOLYHFYDQCEPVIYHNSQDKPKNKY 239  
LN++ +++W+ + + YN IG D YG+ + P+I+HYN++ KFW  
Sbjct: 185 SIINLIFDKWIKLNRGVNYLIGTDYLFYFKTRYLEDGETIPLIHYNTEAKFWLNI 244

Query: 240 SQSKYSSEWYFYGLEWSVIYAQ 262  
+R+R+ +N+Y+ L W IYA+  
Sbjct: 245 FNTFRFNIIWFFYELAWQDIYAK 267

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1306**

A DNA sequence (GBSx1386) was identified in *S. agalactiae* <SEQ ID 4003> which encodes the amino acid sequence <SEQ ID 4004>. Analysis of this protein sequence reveals the following:

Possible site: 13  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2996 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1440-

A related GBS nucleic acid sequence <SEQ ID 10003> which encodes amino acid sequence <SEQ ID 10004> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AACT5095 GB:AB000294 putative Galf transferase [Escherichia coli K12]  
Identities = 68/286 (23%), Positives = 122/286 (41%), Gaps = 18/286 (6%)

Query: 77 STFMGGIAGLRGDIVVQVPTWSTFDELPLDKLQAYGARIITFVHDIVPMFSSNF 136  
S ++ + GL D++P P F +L + RI+ +HDI L

10 Sbjct: 50 SVKLTFFGLGNKDLVLPNFPMAKFWHLSPFRLKKF--RIVPLIHIDELRGSGS 107

Query: 137 YLLDRVIDMYNRSQVVLPTKAMHDYLIKGMITSKVLYQFVNDHPVNIIDLPRPEC---Q 193  
D V D+VI M YL K M+ K+ ++D+ V+ D+ + Q

15 Sbjct: 108 ---DSV--RLATCDMVISHNPQMTIKYL-SKTMQDQIKDKIKIPDYLVSSDVEHRDVTDKQ 161

Query: 194 KVLSPAGDIQRFPFVNDWKENIPLIYYGDSRLNSHANVHAQGMKDQVLMLSLKRQ-G 252  
++ +AG++ R + E +G ++ N G D+ ++ G

20 Sbjct: 162 RGVIIAGNLSRHKCSFIYTEGCDFTLGG--VNTENKDNPKYLG-SFDAQSPKINLPQM 218

Query: 253 FGLCNSEDREELVERR---YSRMAASYKLSTFLAAGLPIANHDISSRDFIKQHLGFTV 309  
FGL W D E Y + N +K S +L+ LP+ + DFI + +G+ V

25 Sbjct: 219 FGLINDGDSVETCSGAPGDLKFNPHKTSIYLSMELPVFINDKAALADPIVDNRIGYAV 278

Query: 310 ETLREAEVKINMEKETYSVVENVEKIALTLLENGYITKLLIDAV 355  
+++E E +++M ETY EN + I+ +R G + +L + +

25 Sbjct: 279 GSIKEMQIVDGMTIETKQISENTKILQIKRTGSYFRDVLVEVI 324

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 30 Example 1307

A DNA sequence (GBSx1387) was identified in *S.agalactiae* <SEQ ID 4005> which encodes the amino acid sequence <SEQ ID 4006>. Analysis of this protein sequence reveals the following:

35 Possible site: 33  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3098 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA73093 GB:M76233 [Rabbit smooth muscle myosin light chain  
kinase mRNA, complete CD6.], gene product [Oryctolagus  
cuniculus]

45 Identities = 23/63 (36%), Positives = 36/63 (56%)

Query: 5 QPAPALQKVRQCCQPAFVLQPVPRCPQALALQVRQCCQPAVLQCVPRCPQPAVLQVPRC 64  
+PA L+ V +PA L+V +PA L+ V +PA+ L+ V +PA+ L+ V

50 Sbjct: 225 KPAETLKFVGNAPASTLKFVGNAPASTLKFVGNAPASTLKFVGNAPASTLKFVGNAPASTLKFVGN 284

Query: 65 QPA 67  
+PA

Sbjct: 285 KPA 287

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1441-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1308**

A DNA sequence (GBSx1388) was identified in *S. agalactiae* <SEQ ID 4007> which encodes the amino acid sequence <SEQ ID 4008>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.24 Transmembrane 189 - 205 (173 - 245)
 INTEGRAL Likelihood = -9.24 Transmembrane 213 - 229 (206 - 245)
10 INTEGRAL Likelihood = -7.96 Transmembrane 95 - 111 (83 - 185)
 INTEGRAL Likelihood = -7.96 Transmembrane 115 - 131 (112 - 185)
 INTEGRAL Likelihood = -7.96 Transmembrane 135 - 151 (132 - 185)
 INTEGRAL Likelihood = -7.96 Transmembrane 155 - 171 (152 - 185)
15 INTEGRAL Likelihood = -6.85 Transmembrane 15 - 31 (8 - 45)
 INTEGRAL Likelihood = -4.09 Transmembrane 39 - 55 (35 - 57)
 INTEGRAL Likelihood = -4.09 Transmembrane 63 - 79 (59 - 81)
 INTEGRAL Likelihood = -2.71 Transmembrane 235 - 251 (235 - 251)
 INTEGRAL Likelihood = -0.11 Transmembrane 253 - 269 (253 - 269)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.
>GP:AC16164 GB:AF010496 ice nucleation protein [Rhodobacter apulensis]
Identities = 85/286 (29%), Positives = 119/286 (40%), Gaps = 17/286 (5%)

Query: 3 ALVLADVDALVETLVLADVDVLEALVLADIEALV---EALVLADIEALVEALVLADID 58
30 AL A AL T + A + L AD+ L +AL A I AL + + A
Sbjct: 523 ALDAQAGALTSTQIGLSTAAVGLSTADMGULTAEQALTSQIALSSSQIRAKTT 582

Query: 59 ALVEALVLADIEALVEALVL---ADIDALVEALVLADVEALIEALVLVEALVLADVE 114
A + AL A I + L + + L ADI AL A + I AL +LV A + AD+
35 Sbjct: 583 AQIALGTAQIKGLTASNIGLETADIVALTTQAPALSSSQIALSTSLVAMETADLA 642

Query: 115 ALIEALVLAL---VEALVLADVEAL---IEALVLVEALVLADVEALIEALVLAVE 166
L A + AL A A+ I + A + L AD+ AL A + +
40 Sbjct: 643 KLSAATFKPSSTQITALTQAAGAITDQIAQITTAIKGLSADIALANATLAKWT 702

Query: 167 ALVLADVEALIEALVLADVD-ALVLALVEALVLVEALIEALVEALVLVEALVLALV 225
A V A + L ++ L A V+AL A + L ++ AL AL V
Sbjct: 703 AQVAVLSSAQLTGLTTQINTVLTAAQVKALGAALAGLTDIVALTGQAALSSQV 762

45 Query: 226 EALILALVEALVLADVDALMEALVLADVEALMEALVLADVDALVEA 271
AL A + AL AD AL A + +AL +DAL A
Sbjct: 763 AALSTAQISALQTADFAALSTAALKGLSSTQITALTSGQIDALTA 808

```

No corresponding DNA sequence was identified in *S. pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1309**

A DNA sequence (GBSx1389) was identified in *S. agalactiae* <SEQ ID 4009> which encodes the amino acid sequence <SEQ ID 4010>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

```

-1442-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2297(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1310**

A DNA sequence (GBSx1390) was identified in *S.agalactiae* <SEQ ID 4011> which encodes the amino acid sequence <SEQ ID 4012>. This protein is predicted to be fimbriae-associated protein Fap1. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3138(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA97453 GB:AB029393 streptococcal hemagglutinin [Streptococcus gordonii]  
 Identities = 388/968 (40%), Positives = 518/968 (53%), Gaps = 68/968 (7%)

Query: 13 VDTKSRVQMKHSEKQWVVTWSEFNLKFAIKGRATVEADVCIQVEKEDRLSSNLITLK 72

V+ +R K+ KS K+R+R S F L + +KG +V V +E + G L YLK

Sbjct: 13 VERVTRFKLINSKGNHLRAATSQFGLRLKMGADISSVEV---KVAESQSVKGGGLNLYK 69

Query: 73 GILAGALVGGASLTSR-VYADETPVQEQSSSVPTLAEQTEVTI-KTTIVQNRQDQTV 129

GI+A GA++GGA +TS VYA+E +++ + LA + E + + T + +

Sbjct: 70 GIIATGAVLGGAVTSSSVYAREEQALEKVILTRDVLATRGCAVLSEAAATLSEGANP 129

Query: 130 SKNIIDGNSVSMSESASTSTSESVSMSMGSTLTSSVESVTSALTSAESISTSASSEV 189

+++ D+ S S S SA+ S S S+S+S S S S S S+S+ S+SES S S S SV

Sbjct: 130 VESLEDTLSASESASAN-SVSTISISSESFSVSASASLESSSSELSQSSSESASASISLV 188

Query: 190 SKSTSISEVSNILSTQASLIDNGKSPSQAQIVTESSLVTDAGKNSVSLIRITKPKSE 249

S S TS S S TQ+S + S S+N + T S V+ +NA V + + +E

Sbjct: 189 SASTSQSPSTTSSTQSSNNBSLSSDSSNLNTNQS-VSARQNRVETRRVAANLTS 247

Query: 250 LQTSKMSNLSLTPEKSQWMIASDKTGNESLTPTIRLKSIVIQPRSMNLTSSEMDLPL 309

K + + E + ++ TN ++ ++ N+ ++ L P

Sbjct: 248 APQVKSQDVYVYRGSPEFY--AKITDNGQGVNRVIR-----NVDGANSYLGEPN 297

Query: 310 EEVSPTMLGKDVSSSLQKVNIADKNTLSSEPGTVKLDSSENVLNLPFASIASVNGSIVF 369

TE LG+ ++ +Q L+ E ++ ++ + +A G+

Sbjct: 298 WVKYSTENLGRPGNATVQN---PLRTRIPGEVPIANEIVNEKSYTRYI--VAMPQSGN-- 350

Query: 370 TVKLSNLDTCQIGTILKQVQIDMETQQLLATGSYSPLAHNITY-----TWIRYAST 421

++ DN + G+ + +E Y P ++TY T R A

Sbjct: 351 ATQWVDNANRNGLERFVLITVKSQNE-----KYDPKSSVITYNNLSLSTSEREAVA 402

Query: 422 LNNIKARVNPVWPDQRI-----ISKITSDKQCTATLANQVASIE---ERVQYNSP 471

A N+P P +I ++ T DK T N V ++ S S

Sbjct: 403 AAVRAANPNIP--PTAKITVSQGTVTITYPKSTDTIPANRVKDLQISKNSASQSS 460

Query: 472 VTEHNVKTNVRSIMKLDDEQETIYITQINPEGKMYFASGLNLYITIGSDGTSGPS 531



-1443-

```

 V+ + T+V + I ++ + + ++ S+ S S
Sbjct: 461 VSAQSQASSTSVASBI---SASMSASVSSTASTSASVSASBSASTSASVSASBSASTS- 516

5 Query: 532 VNLLNAEVLKIKTNSKNLIDSMQNYDSPEFEDVTSQYSYTMDSKITIDWKTNISISSTT 591
 A V K++S + + S ++ + + S + S + S+S+T
Sbjct: 517 -----ASVSASKSSSTASVSASBSASTSASVSASBSASTSASVSASBSASTSASVSAST 571

 Query: 592 SYVVLVKIPKQGVLYSTVSDINQTYGSKYSYGHINISGDSANARIKL-LSEASTSAS 650
 S + ST + ++ + + S ++S A+ + SEASTSAS
10 Sbjct: 572 SASTSASVSASBSA--STASVSASBSASTS---ASVASEASTSASVSASBSASTSAS 626

 Query: 651 TSASTSASMSASTSASTSASMSASTSASTSASTSASMSASTSASTSASTSASTSASTSAS 710
 SAS S+S SAS SAS SAS SAS SAS SASTSAS+SASTSASTSAS SASTSASTSAS
15 Sbjct: 627 VBASEBSSTASVSASBSASTSASVSASBSASTSASVSASTSASTSASVSASTSASTSAS 686

 Query: 711 MSASTSASTSASTSASTSASTSASTSASMSASTSASTSASTSASTSASTSASTSAS 770
 +SASTSASTSAS SAS SASTSAS SAS SASTSAS SASTSASTSAS+SASTSASTSAS
Sbjct: 687 VASTSASTSASVSASBSASTSASVSASBSASTSASVSASTSASTSASVSASTSASTSAS 746

20 Query: 771 TSASTSASMSASTSASTSASTSASTSASMSASTSASTSASTSASTSASTSASTSAS 830
 SAS SAS SAS SASTSASTSAS SAS SASTSAS SASTSASTSAS+SASTSASTSAS
Sbjct: 747 VBASESASTSASVSASTSASTSASVSASESASTSASVSASTSASTSASVSASESASTSAS 806

 Query: 831 TSASMSASTSASTSASTSASTSASTSASMSASTSASTSASTSASTSASTSASTSAS 890
 SAS SASTSAS SAS SASTSAS SAS SASTSAS SASTSASTSAS+SASTSASTSAS
25 Sbjct: 807 VBASESASTSASVSASTSASTSASTSASVSASESASTSASVSASESASTSASVSASESASTSAS 866

 Query: 891 NSATTEASTSVSTASTASTSASTSSSSSVTNSKSKVYSALEFSTGDQDYSVTATLQ 950
 +S+TSASTS S SAS SASTSAS S+S S ++++S SA S +T+
30 Sbjct: 867 VASTSASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTSAS 926

 Query: 951 LGLMTGAT 958
 + T A+
Sbjct: 927 VASTSAS 934

```

There is also homology to SEQ ID 760.

SEQ ID 4012 (GBS68) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 4; MW 131.2kDa).

GBS68d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 14; MW 103kDa) and in Figure 239 (lane 13; MW 103kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 17; MW 78kDa), in Figure 153 (lane 17; MW >78kDa) and in Figure 184 (lane 10; MW 78kDa). Purified GBS68d-GST is shown in Figure 246, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1311

A DNA sequence (GBSx1391) was identified in *S.agalactiae* <SEQ ID 4013> which encodes the amino acid sequence <SEQ ID 4014>. This protein is predicted to be RofA. Analysis of this protein sequence reveals the following:

```

50 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1738 (Affirmative) < succ>
55 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-1444-

A related GBS nucleic acid sequence <SEQ ID 10005> which encodes amino acid sequence <SEQ ID 10006> was also identified.

There is also homology to SEQ ID 3750.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1312

A DNA sequence (GBSx1392) was identified in *S.agalactiae* <SEQ ID 4015> which encodes the amino acid sequence <SEQ ID 4016>. This protein is predicted to be Nra. Analysis of this protein sequence

10 reveals the following:

```
Possible site: 16
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

There is also homology to SEQ ID 3750.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1313

A DNA sequence (GBSx1393) was identified in *S.agalactiae* <SEQ ID 4017> which encodes the amino acid sequence <SEQ ID 4018>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3674 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA27020 GB:M80215 uvs402 protein [Streptococcus pneumoniae]
Identities = 577/663 (87%), Positives = 633/663 (95%), Gaps = 1/663 (0%)

Query: 1 MIDRKDTNRFKLVSKYSPSGDQQAIRTLVDNIEGGEKAQILKGATGTGKTYTMSQVIAQ 60
MI+ N+FKLVSKY PSGDQQAIR LVDNIEGGEKAQIL GATGTGKTYTMSQVI++
Sbjct: 7 MINHITDMQFKLVSKYQPSGDQQAIRTLVDNIEGGEKAQILANGATGTGKTYTMSQVISK 66

Query: 61 VNKP TLVIAHNKTLAQLYGEPEKFFPPMAVEYFVSYDDYQPEAYVPSSDTYIEKDSSV 120
VNKP TLVIAHNKTLAQLYGEPEKFFPPMAVEYFVSYDDYQPEAYVPSSDTYIEKDSSV
Sbjct: 67 VNKP TLVIAHNKTLAQLYGEPEKFFPPMAVEYFVSYDDYQPEAYVPSSDTYIEKDSSV 126

Query: 121 NDEIDKLHRSATSSLLERNDIVVASVSCYGLGSPKRYADSVVSLRPGQIRSRDQLLN 180
NDEIDKLHRSATSSLLERNDIVVASVSCYGLGSPKRYADSVVSLRPGQIRSRDQLLN+
Sbjct: 127 NDEIDKLHRSATSSLLERNDIVVASVSCYGLGSPKRYADSVVSLRPGLEISRDKLND 186

Query: 181 LVDIQFERNDIDPQGRGKFRVGRGVVEVFPASRDHAFRIEFGDEIDRIRIEISLGRVL 240
LVDIQFERNDIDPQGRG+FKVRGDEVVE+FPASRDHAFRIEFGDEIDRIRIE+E+LTG+VL
Sbjct: 187 LVDIQFERNDIDPQGRGKFRVGRGVVEVFPASRDHAFRIEFGDEIDRIRIEVRLTGQVL 246
```

-1445-

Query: 241 GEVHLAIFPAATHFMTNDHMEEAISKIQAEMNQVELFEKSGKLEAQRIRQRTYDIIR 300  
 GEV+HLAIFPAATHF+TND+HME AI+KIQAEE+Q+ +FEKSGKL+EAQR++QRTYDIIR

Sbjct: 247 GEVDHLAIFPAATHFVTDNDHMEVAIAKIQALSEQLAVFEKSGKLEAQRIRQRTYDIIR 306

5 Query: 301 MLREMGYINGVENYSRHMGRSGEGPPPTLLDFFPDFLIMIDESHMTMGQIKSMYNGDR 360  
 MLREMGYINGVENYSRHMGRSGEGPP+TLLDFFP+DFLIMIDESHMTMGQIKSMYNGDR

Sbjct: 307 MLREMGYINGVENYSRHMGRSGEGPPPTLLDFFPDFFDPLIMIDESHMTMGQIKSMYNGDR 366

10 Query: 361 SRKEMLVNYGFRPLPSALDNKPLRKEEFESHVHQIVVYSATPGDYEMBGQITVTVBQIIRPT 420  
 SRK+MLVNYGFRPLPSALDNKPLRKEEFESHVHQIVVYSATPGDYH BQT+TV+BQIIRPT

Sbjct: 367 SRKQMLVNYGFRPLPSALDNKPLRKEEFESHVHQIVVYSATPGDYENBQITVTVBQIIRPT 426

Query: 421 GLLDPEVEVRPSMGQDMLLGEINLRTEKGERFTITTLTKMAEDHLDYLEKMGVVKVM 480  
 GLLDPEVEVRK+MGQ+DMLLGEINLR E K ERTITTLTK+MAELTDY KMG+VKVM

15 Sbjct: 427 GLLDPEVEVRPTMGQDMLLGEINLRVKNRFTITTLTKMAEDHLDYLEKMGVVKVM 486

Query: 481 HSDIKTLERTEIIRDLRGVDFVLGINLLRBGIDVFEVSLVAILEDADKGFLENGRLI 540  
 HSDIKTLERTEIIRDLRGVDFVL+GINLLRBGIDVFEVSLVAILEDADKGFLENGRLI

20 Sbjct: 487 HSDIKTLERTEIIRDLRGVDFVLGINLLRBGIDVFEVSLVAILEDADKGFLENGRLI 546

Query: 541 QTIGRAARNNGHVIHYADKITSQRADETAARRRLQMDVNEKGIIVPQTIKKGIIRDL 600  
 QTIGRAARNS GHVIHYAD +T SMQRA+DETAARR++QM YNE+HIVPQTIKKGIIRDL

Sbjct: 547 QTIGRAARNSEGHVIHYADFTVTSQRAIDETAARRRKIQMAYNEENGIVPQTIKKGIIRDL 606

25 Query: 601 IAITKENDSKPEKVVYSSLSKIKERQARIKALQQQQAELLDPELAAGIRDVILEK 660  
 IATK+ ++ +K VD +SLA+K+ER+ +K L+QMQEA E+LDPELAQIRDL+LE+K

Sbjct: 607 IAVTKAVAKEE-DKEVDINSINKERKEIVKLEKQMQAENVVLDPELAQIRDMLEVK 665

Query: 661 AID 663  
 A+D

30 Sbjct: 666 ALD 668

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4019> which encodes the amino acid sequence <SEQ ID 4020>. Analysis of this protein sequence reveals the following:

35 Possible site: 55  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4386 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 570/663 (85%), Positives = 625/663 (93%)

45 Query: 1 MIDRKIDNPKFLVSKYSPSGDQPOAIFLWNIIEGGEKAQILKGATGTGKTYTMSQVIAQ 60  
 MID+D PKL SKY PSQGDQQAIE+LWNIIEGGEKAQIL GATGTGKTYTMSQVIA++

Sbjct: 1 MIDKRDDEKPFKLSKYKPSQGDQQAIESLWNIIEGGEKAQILGATGTGKTYTMSQVISK 60

50 Query: 61 VNKPTLVIAHNKTLAQQLYGEPEKFFPDNAVYFVSYDYDYQPEAVPSSDYIEKDSV 120  
 VNKPTLVIAHNKTLAQQLYGEPEKFFPDNAVYFVSYDYDYQPEAVPSSDYIEKDSV

Sbjct: 61 VNKPTLVIAHNKTLAQQLYGEPEKFFPDNAVYFVSYDYDYQPEAVPSSDYIEKDSV 120

55 Query: 121 NDEIDKLRHSATSSLERNDVIVVASVSCYIGLSPKETAADVSLPQGEISRDILNQ 180  
 NDEIDKLRHSATSSLERNDVIVVASVSCYIGLSPKETAADVSLPQGEISRD ILNQ

Sbjct: 121 NDEIDKLRHSATSSLERNDVIVVASVSCYIGLSPKETAADVSLPQGEISRDILNQ 180

Query: 181 IVDIQFERNDIDPQRCFVRVGDVVEVFPASRDEHAFRVEFGDIDRIDRESIAGRLV 240  
 IVDIQFERNDIDPQRC FVRVGDVVEVFPASRDEHAFR+EFPGDIDRIDR EIESLAG+ +

60 Sbjct: 181 IVDIQFERNDIDPQRCFVRVGDVVEVFPASRDEHAFRVEFGDIDRIDRESIAGKVI 240

Query: 241 GEVHLAIFPAATHFMTNDHMEEAISKIQAEMNQVELFEKSGKLEAQRIRQRTYDIIR 300  
 GEV+HL +PAAATHF+TNDHME++I+KIQAEE+ Q++LFE ECKL+EAQR+QRTYDIIR

Sbjct: 241 GEVDHLAIFPAATHFMTNDHMEQAIAKIQALFESEKGLLEAQRIRQRTYDIIR 300

-1446-

Query: 301 MLRENGYTGNGVENYSRHMDSRSGEPPPTLLDFFPEDFLINIDESHMTWQIKMYNDR 360  
MLRENGYT+GVENYSRHMDSRSGEPPPTLLDFFPEDFLINIDESHMTWQIKMYNDR+

Sbjct: 301 MLRENGYTGNGVENYSRHMDSRSGEPPPTLLDFFPEDFLINIDESHMTWQIKMYNDR 360

5 Query: 361 SRKEMLVNYGFRPLPSALDNRLRKEEPFESHVHQIYVVSATPGDYEMOTDVTVVBQIIRPT 420  
+RK+MLV+YGFRPLPSALDNRLRKEEPFESHVHQIYVVSATPG+YEM QT+T++BQIIRPT

Sbjct: 361 ARKQMLVDYGFRLPSALDNRLRKEEPFESHVHQIYVVSATPGYEMQVTVTITIEQIRGL 420

10 Query: 421 GLLDPSVEVRPSMGMDDLLGEINLRTEKGERTFTITLTKRMAEDLDYLKEMGVKVKYK 480  
GLLDPE++VR SMGMDDLLGEIN R + ERTFTITLTK+MAEDLDYLKEMGVKVKYK

Sbjct: 421 GLLDPEIDVSSVMGMDDLLGEINQRVARDERTFTITLTKKMAEDLDYLKEMGVKVKYK 480

Query: 481 HSDIKTLERTEIIRDRLGVDFDLGINLLRSGIDVPEVSLVAILDADKRGFLNNGRLI 540  
HSDIKTLERTEIIRDRLGVDFDLGINLLRSGIDVPEVSLVAILDADKRGFLNNGRLI

15 Sbjct: 481 HSDIKTLERTEIIRDRLGVDFDLGINLLRSGIDVPEVSLVAILDADKRGFLNNGRLI 540

Query: 541 QTIGRAARNNGHVIMYADKTTDSMQRANDSTARRELLQMDYNEKMGIVPOTIKKEIRD 600  
QTIGRAARN +GHVIMYADK+TDSMQPA-DETAARR +Q+ YN+ HGIVPOTIKK+IR L

Sbjct: 541 QTIGRAARNVDGHVIMYADKTTDSMQRANDSTARRELLQIAYNKGMSVTPOTIKKEIRD 600

20 Query: 601 IAITKSNDSKPEKVVDDYSSLSKKERQAEIKALQQQMGAELLDPELAAQIRVILELK 660  
I-I+K++ +D ++ +DY S+S+ ER+ I ALQ+QMGAELLDPELAAQ+RD+ILELK

Sbjct: 601 ISIKTSKNDISKEMDYSSMSGRKKEATNALQKQMGAELLDPELAAQIRVILELK 660

25 Query: 661 AID 663  
+D

Sbjct: 661 LMD 663

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1314

A DNA sequence (GBSx1394) was identified in *S. agalactiae* <SEQ ID 4021> which encodes the amino acid sequence <SEQ ID 4022>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

| INTEGRAL | Likelihood          | Transmembrane | 284 - 300 ( 274 - 303) |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -11.78 | Transmembrane | 20 - 36 ( 16 - 53)     |
| INTEGRAL | Likelihood = -10.08 | Transmembrane | 117 - 133 ( 114 - 137) |
| INTEGRAL | Likelihood = -5.52  | Transmembrane | 203 - 219 ( 201 - 225) |
| INTEGRAL | Likelihood = -5.15  | Transmembrane | 183 - 199 ( 182 - 200) |
| INTEGRAL | Likelihood = -3.29  | Transmembrane | 74 - 90 ( 73 - 90)     |
| INTEGRAL | Likelihood = -1.54  | Transmembrane | 37 - 53 ( 37 - 53)     |
| INTEGRAL | Likelihood = -0.48  | Transmembrane |                        |

----- Final Results -----

| bacterial membrane  | --- Certainty=0.5713 (Affirmative) < succ> |
|---------------------|--------------------------------------------|
| bacterial outside   | --- Certainty=0.0000 (Not Clear) < succ>   |
| bacterial cytoplasm | --- Certainty=0.0000 (Not Clear) < succ>   |

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:CAA22372 GB:AL034446 putative transmembrane protein  
[Streptomyces coelicolor A3(2)]  
Identities = 58/190 (30%), Positives = 96/190 (50%), Gaps = 11/190 (5%)

Query: 114 GWS--IQFILFSVITAYILGLDPSHYDVK-ATIFVVVFLPFWLQSSTELLIRG 170  
GW IQF LF +VIT G Y+V ++ + L+ F + TEE++ R

55 Sbjct: 98 GWSLILGFGLPG-AVITNLFSQY----YEVDLGVSQGAIGLVGFMAAAATEEVFRG 152

Query: 171 WLLPLINIRPHLAVAIGVSSTLFGLHLVNAHVTLFSIVSI-ICSGVMSLYMKSGNIW 229  
L +I +A+G++ +FG+HL+N T ++I +G+++ +N+W

60 Sbjct: 153 VLFRIIEEHGTYLALGLTGLVFGMLHLNEDATLWALATAIEAGPLAAAYATNRLM 212

Query: 230 SVAALLIGAWNVSQNLVGIASVGSGKAGASLLHFTVKENAPWISOGAGIEGSLISIFVL 289

-1447-

+H WNF+ G ++ VSG LL T+ + P ++GG FG EGS+ B+  
 Sbjct: 213 LTIGVHFGWFWAAGGVFSTVVSQKQDSBKLLDQTN--SGFKLLTGDDFGFEGSVSVSGFG 270  
 Query: 290 LAATIIYLLML 299  
 + + LML  
 Sbjct: 271 VLITLVFLML 280

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1315

A DNA sequence (GBSx1395) was identified in *S.agalactiae* <SEQ ID 4023> which encodes the amino acid sequence <SEQ ID 4024>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -8.97 Transmembrane 532 - 548 ( 523 - 553)  
 INTEGRAL Likelihood = -7.38 Transmembrane 700 - 716 ( 696 - 720)  
 INTEGRAL Likelihood = -4.57 Transmembrane 562 - 578 ( 558 - 588)  
 INTEGRAL Likelihood = -0.32 Transmembrane 665 - 681 ( 665 - 681)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4589 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF16724 GB:AF141644 putative integral membrane protein  
 [Lactococcus lactis]  
 Identities = 109/195 (55%), Positives = 156/195 (79%), Gaps = 4/195 (2%)  
 Query: 466 MFPNGLASLKKSGEYDKLVKKYLSTASTSSNDKAAKPDVESTILGLISNNYKQLLSGIG 525  
 +MFPNGLA+L+ +GEYDK++ KYL++ T + +AK E+T G++ NN++Q+ G+  
 Sbjct: 1 EMFPNGLANLRANGEYDKIIDKYLAS-DTKTIQSSAK---ENITFFGLQNNWEGIQIGSL 56  
 Query: 526 TTLSITLISPAIAMVIGIIPGMMSVSPNLTRTISMIFDIVIRGIPLMIVAIFWIGIPW 585  
 TL L ++SF +AM++GIIFG+ SV+PS LRTI+ I+VD+ R IPL+++ FIF+GIPN  
 Sbjct: 57 VTLEAVLSFILAMIVGIIIFGLFSVAPSKILRTIARIYVDLNRSLPLLVLITIFIFYGIPN 116  
 Query: 586 LIBSITGHSQSPINDFVAATIALSLNGAYIARIVRGGIBAVPSGQMRASRLSGYSGKTN 645  
 L++ ITGHSQSP+N+F A IAL+LN AYIARIVR G++AVPSGQMRASRLG++Y +H  
 Sbjct: 117 LQIITGHSQSPINFTGVALITNSGAYIARIVRSGVQAVPSGQMRASRLGVTILTSN 176  
 Query: 646 QKVILFQAVRIMLPN 660  
 +KVILFQ+ ++ P+  
 Sbjct: 177 RKVILFQAIKITIPS 191

There is also homology to SEQ ID 1198.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9071> which encodes amino acid sequence <SEQ ID 9072>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> May be a lipoprotein  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1448-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS sequences follows:

Score = 80.8 bits (196), Expect = 2e-17  
 Identities = 64/233 (27%), Positives = 113/233 (48%), Gaps = 13/233 (5%)

Query: 34 IKKTRKLVAVSPDYAPFPEFKALVNGKDTIVGADVQLAQAIADLEVDLELSPMSFDNLV 93  
 +K + K+V S +APPE+ NGK G D+L + IA + L++S FD L  
 Sbjct: 268 VKPSYKIVSDSS--FAPPEYQ---NGKGYTFDFMELIKLIAQQGFKLDISNPGFDNAL 322

Query: 94 SSLTGKADIATISGISHTKERAKVYDFSIPYQARNAIVMRASDAKVTKNISDLNGKKVA 153  
 +++Q+G+AD I+G + T+ R K++DFS PYY +++++ K+ DL GK V  
 Sbjct: 323 NAVQSGQADGVLAGATTTEARQKLFPSDPY--TSSVILAVKKGSNVKSYQLKSKTIVG 380

Query: 154 AOKGSIEEGLVKIQLPKANLISLTANGEA---INELKAGQVYAVTLEAFVAGFLAQHKD 210  
 A+ G+ + K N + A EA + + +G + A+ + V A + Q +  
 Sbjct: 381 AKNGTASYIVLSDHADKYN-YHVAFDEASTMYDSMNSGSTDALMDDEAVLAYALNQGRK 439

Query: 211 LALAPFSLKTSQDGAKAVALPKNSGDLTKAVNKVIAKLDBQERYKSFIAETIA 263  
 P + S GD + +L K N +A L + Y + + + +  
 Sbjct: 440 FR-TPIKGERS-GDIGFAVKGANPELIKMFNNGLASLKSGEYDKLVKKYLS 490  
 Score = 74.5 bits (180), Expect = 1e-15  
 Identities = 59/215 (27%), Positives = 102/215 (47%), Gaps = 12/215 (5%)

Query: 48 YAPFEFKALVNGKDTIVGADVQLAQAIADLEVDLELSPMSFDNLVSSLTGKADIATISG 107  
 YAPFEFK + T G DV + +A ++ + + FD +++Q+G+AD +G  
 Sbjct: 36 YAPFEFK---DSDQTYKGIDVDIVNEVAKRAGNVNNTTPGFDAAVNAVQSGQADALMAG 92

Query: 108 ISHTKERAKVYDFSIPYQARNAIVMRASDAKVTKNISDLNGKKVAOKGSIEEGLVKIQ 167  
 + T+ R KV+PS YY + I+ ++ KVT N L GK V + G+ + +  
 Sbjct: 93 TTVTEARKKVFNFSDTYDT-SVILYTKNNKVT-NYQLKGVGVGVNGSTAAGSLEEN 150

Query: 168 LPKANLISLTANGEAI--NELKAGQVYAVTLEAFVAGFLAQHKDLALAPFSLKTSQDGA 225  
 K T + N L +G +YA + FV + Q K A+ + + + +  
 Sbjct: 151 KSKYGYKVTFTDSDLGNLSLDSGSIYAAMDQFVQFALNQSKAYAI---IMMGEAVGS 207

Query: 226 KAVALPKNSG--DLTKAVNKVIAKLDBQERYKSF 258  
 A A+ K SG +L K N A++ Y +  
 Sbjct: 208 FAFVAKKSGSDNLKIFENFAQMKSDGTYNIM 242

SEQ ID 4024 (GBS154) was expressed in *E. coli* as a His-fusion product. The purified protein is shown in Figure 199, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1316

A DNA sequence (GBSx1396) was identified in *S. agalactiae* <SEQ ID 4025> which encodes the amino acid sequence <SEQ ID 4026>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein (glnQ). Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4183(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB90561 GB:AB001058 glutamine ABC transporter, ATP-binding

-1449-

protein (glnQ) [Archaeoglobus fulgidus]  
Identities = 147/240 (61%), Positives = 192/240 (79%)

- 5 Query: 5 KIDVDLHKSYGQNEVLKIGIDAKFYEGDVVCIIGPSGSGSKSTFLRTLNLESITSKVVV 64  
++++ DLHK +G+ EVLKG+ K +G+VV IIGPSGSGSKET LR +N LE TSGK+++  
Sbjct: 3 QLEHIDLHKRFGELEVLKGVTKRVEKGVVVIIGPSGSGSKSTFLRCINRLSPSTSGKILL 62
- 10 Query: 65 DGFELSNPKTDIDKARENIGMVPQHNLPPHMSVLENTIAPTELKESKEAAEKHGKEL 124  
DG +++N K DI+K R+ IG+VPQ FNLFPH++ L+N+T API++ K SK AB+ GM L  
Sbjct: 63 DGVDTINSKIDINKVQRIGIVFQQNLFPHLTAHQVTLAPTKIKMSKREABELGRL 122
- 15 Query: 125 LEKVGGLADKANAKPDSLGGGQKQVAIARSLAMNPDIILFDEPTSLDPEMVGDLVNMK 184  
LEKVGEL DKA+ P LSGGQ+QKVAIAR+LAMNP++LDFE TSLDPE+V +VL+VMK  
Sbjct: 123 LEKVGELDKADYYPALSGGQQKQVAIARALAMNPEVMLFUEVTSALDPELVKVLDMK 182
- 20 Query: 165 DLABQGMIMLVITHMGMFARQVNRVIFPDGGRFLDEGTFEQIPDTPQHPRQDPLNKVL 244  
LA GMTM++VTHMGMFAR+V+RVIF DGG +B+G PEQIF P+H R + FL+ +L  
Sbjct: 163 QLARDGMIMVVVTHMGMFAREVGRVIFPDGGVTVBEGKPEQIFSNPKHERKTRKFLSMIL 242
- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4027> which encodes the amino acid sequence <SEQ ID 4028>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.4149 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has homology with the following sequences in the databases:

>GP:BAB05180 GB:AP001512 ABC transporter (substrate-binding protein)  
[Bacillus halodurans]  
Identities = 79/227 (34%), Positives = 126/227 (54%), Gaps = 10/227 (4%)

35 Query: 35 FKTRKLIVAVSPDYAPFEKALVNGKDTIVGADVQLAQAIADLELDVLELSPMSFANVL 94  
+K LV+ S DY P+E + G+ IVG DV +A+ I EL +L++ M P+ ++  
Sbjct: 48 EKKSVLVMGTSADYPPYESSVDVTGIR--IVGFDVDAEYITSELGYELKIQDMDFGIITP 105

40 Query: 95 SLTGKADLAISGISHTKERAKVYDFSIPIYQAENAIVMRASDAKVTKNISDLNKKVAA 154  
+LQ G+ D A+SG++ T+ER K DFS YY A+H +V + D ++ DL GK V  
Sbjct: 106 ALQAGRVDFALSGMVTPEERKKSVDFSDVYDAQNLFVVKBEDG--LSSVEDLAGTGVGV 163

45 Query: 155 QKQSI--EEGLVKIQ--LPHANLISLPMGEAINELKAGQVAVTLEAFVAAGFLAHKDL 211  
Q SI EE V++Q L + + + E + EL AG+V A+ +E VAAG L + +  
Sbjct: 164 QLASIQGEAAVELQRELOGLTETERNRVVELVQELLAGRVDAIITEDTVAAHLEANT-- 221

50 Query: 212 ALAPPSLKTSDGDAKVALPKNSCDLTKAVNKVIAKLDQBRYSKSI 256  
L P++++ A+A FK+S +LT+ N+ ++ + + I  
Sbjct: 222 GLVRFAIESBEGTGSALFAPKDS-ELTEPFNEKIQMMEDGTMEELI 267

An alignment of the GAS and GBS proteins is shown below.

Identities = 223/246 (90%), Positives = 236/246 (96%)

- 55 Query: 1 MHELKIDVDLHKSYGQNEVLKIGIDAKFYEGDVVCIIGPSGSGSKSTFLRTLNLESITS 60  
M ELKIDVDLHKSYGQNEVLKIGIDAKFYEGDVVCIIGPSGSGSKSTFLRTLNLE+ITSG  
Sbjct: 1 MTELKIDVDLHKSYGQNEVLKIGIDAKFYEGDVVCIIGPSGSGSKSTFLRTLNLETTITSG 60
- 60 Query: 61 KVVVDGFELSNPKTDIDKARENIGMVPQHNLPPHMSVLENTIAPTELKESKEAAEKH 120  
KV+VDGFELS+PKT+IDKARENIGMVPQHNLPPH+VLENI PAP+ELGKSKE A+KH  
Sbjct: 61 KVVVDGFELSDPKINIDKARENIGMVPQHNLPPHMTVLENTIAPAVELGKSKEVAKKH 120
- Query: 121 GMELEKVGGLADKANAKPDSLGGGQKQVAIARSLAMNPDIILFDEPTSLDPEMVGDL 180  
GM LLEKVLG+DKA+A P LSGGQKQKVAIARSLAMNPDI+LFDPTSLDPEMVGDL  
Sbjct: 121 GMALEKVGLESDKANAPPSLGGGQKQVAIARSLAMNPDIILFDEPTSLDPEMVGDL 180

-1450-

Query: 181 NVMKDLAEQGNTMLIVTHEMGPARQVANRVIFTDGGRFLEDGTFEQIDTFQHPRLQDFL 240  
 NVMKDLAEQGNTMLIVTHEMGPARQVANRVIFTDGG+FLDGGTFE+IFD P+HPRL +FL  
 Sbjet: 181 NVMKDLAEQGNTMLIVTHEMGPARQVANRVIFTDGGRFLEDGTFEQIDTFQHPRLIEFL 240

Query: 241 NKVLNV 246  
 +KVLNV  
 Sbjet: 241 DKVLNV 246

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1317

A DNA sequence (GBSx1397) was identified in *Sagalactiae* <SEQ ID 4029> which encodes the amino acid sequence <SEQ ID 4030>. Analysis of this protein sequence reveals the following:

15 Possible site: 18  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.2311 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4031> which encodes the amino acid sequence <SEQ ID 4032>. Analysis of this protein sequence reveals the following:

25 Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2702 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 45/51 (88%), Positives = 49/51 (95%)

Query: 1 MGDKPISPRDKDGNFVSADVWNAEKLEELNLTINPNRKLRLEREKLAKK 51  
 MGDKPIS+DKDGNFVSADVWNAEKLEELNLTINPNR+LRLEREKL K++  
 Sbjet: 11 MGDKPISPKDKDGNFVSADVWNAEKLEELNLTINPNRKLRLEREKLKDE 61

- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1318

- A DNA sequence (GBSx1398) was identified in *Sagalactiae* <SEQ ID 4033> which encodes the amino acid sequence <SEQ ID 4034>. This protein is predicted to be spo0b-associated GTP-binding protein (obg). Analysis of this protein sequence reveals the following:

50 Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2967 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>



-1451-

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB14752 GB: Z99118 GTPase activity [Bacillus subtilis]  
Identities = 297/435 (68%), Positives = 345/435 (79%), Gaps = 7/435 (1%)

5 Query: 3 MFLDTAKISVQAGRGGDGMVAFRRKRYVPGGFWGGDGGKGGSVIFKVNGLRLTMDPRY 62  
MF+D K+ VK G GG+GMVAFRRKRYVP GGP GGDGGKGG V+P+V+EGRLTMDPRY  
Sbjct: 1 MFVDQVKVYVKGDSGNGMVAFRKRYVPKGGPAGSGDGGKGGDVVPRVDRGLRLTMDPRY 60

10 Query: 63 NRNPKAKAGEKGMTKGMHGRGARDLIVSLPPGTTVRDATTGRKVTDLVEHQEPVVVARGS 122  
++PKA GE GM+K HGR A+D+++ +PGT V D T +VI DL EH Q V+ARGG  
Sbjct: 61 KKHFKAIRGRHGMKNGHGRNADDMVVKVPPGTTVTDITQVADI/TEGQRAVIRARG 120

15 Query: 123 RGGGRGNIRFATPAPNPAPEIAENGEPGGERELQLELKILADVGLVGPFSVGKSTLLSVSA 182  
RGGGRGN RFATP NPAP+++ENGEPG+ER +LELK+LADVGLVGPFSVGKSTLLSVS+  
Sbjct: 121 RGGGRGNRFATPAPNPAQLSENGEPGKERYIVLELKVLADVGLVGPFSVGKSTLLSVSS 180

20 Query: 183 AKPKIGAYHFTTIVNGLMVRTKSGDSFAMADLEGLIEGASQGVGLTQPLRIHIERTRVI 242  
AKPKI YHFTT+VPLGLMV T G SF MADLPGLEIGA QVGLG QPLRIHIERTRVI  
Sbjct: 181 AKPKIADHYFTT+VPLGLMVRTDGGRSF+MADLPGLEIGAQQVGLTQPLRIHIERTRVI 240

25 Query: 243 LHVIMDSASBGRDPYDDYVSINNELETYNLRIMERPQIIIVANKMMPDSEENLAAPKEKL 302  
+HVIMDS BGRDPYDDY++IN EL YNLR ERQIIIVANKMMP++ ENL APKEKL  
Sbjct: 241 VHVIMDSBGRDPYDDYLTINQLESYNLR/TERPQIIIVANKMMPAAENLAAPKEKL 300

30 Query: 303 AANYDEPDDMPMIFFISSLAHQLENLMDATAELLANTTEFLYDSTDMQDEAYYGNNE 362  
DD P +PETS++ +GL L+ A L MT EF LYDE ++ ++ Y  
Sbjct: 301 T-----DDYP-VFPIAVTRBSLELLFEVANQLENT/EFPLYDEBELQNRVMVTMEN 353

35 Query: 363 DERPFETTRDDATWLYLSDKLEKLFVMTNMRDESIMKFAQLRGHGVDEALREKGA 422  
+B PF ITRD D +VL GD LE+LF MT+ RDES+ +FAQ+RSGHGVDEALREKGA  
Sbjct: 354 EEVFNTRITRDPDGVFLGSDSLEKLFMTNDFSRDESVKRFAQGRHGVDEALREKGA 413

40 Query: 423 GDIVRIGNFEFEFD 437  
GDI+R+ FEFEFD  
Sbjct: 414 GDIIRLLEFEFEFD 428

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4035> which encodes the amino acid sequence <SEQ ID 4036>. Analysis of this protein sequence reveals the following:

40 possible site: 14  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2588 (Affirmative) < succ>  
45 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 394/437 (90%), Positives = 421/437 (96%)

50 Query: 1 MSVFLDTAKISVQAGRGGDGMVAFRRKRYVPGGFWGGDGGKGGSVIFKVNGLRLTMDPR 60  
MSVFLDTAKISVQAGRGGDGMVAFRRKRYVPGGFWGGDGGKGGSVIFKVNGLRLTMDPR  
Sbjct: 1 MSVFLDTAKISVQAGRGGDGMVAFRRKRYVPGGFWGGDGGKGGSVIFRVDGLRLTMDPR 60

55 Query: 61 RYRNPKAKAGEKGMTKGMHGRGARDLIVSLPPGTTVRDATTGRKVTDLVEHQEPVVVAR 120  
RYRN PKAK+GSKGMTKGMHGRGARDLIV +P GTTVRDA TKVTDLVEH GE V+A+  
Sbjct: 61 RYRNPKAKAGEKGMTKGMHGRGARDLIVFPQGTTVRDAETGKVTDLVEHQEPVIAK 120

60 Query: 121 GGGGRGNIRFATPAPNPAPEIAENGEPGGERELQLELKILADVGLVGPFSVGKSTLLSV 180  
GGGGRGNIRFATPAPNPAPEIAENGEPGGER+L+LELKILADVGLVGPFSVGKSTLLSV  
Sbjct: 121 GGGGRGNIRFATPAPNPAPEIAENGEPGGERQLELKILADVGLVGPFSVGKSTLLSV 180

Query: 181 SAAKPKIGAYHFTTIVNGLMVRTKSGDSFAMADLPGLEIGASQGVGLTQPLRIHIERTR 240  
S+AKPKIGAYHFTTIVNGLMVRTKSGDSFAMADLPGLEIGASQGVGLTQPLRIHIERTR

-1452-

Sbjct: 181 SSAXPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQVGLGTQFLRIHIER 240  
 Query: 241 VILHVIDMSASEGRDPYDDYVSINNLETTYNLRIHERPQIIIVANKMDPDSSENLAAPE 300  
 VILHVIDMSASEGRDPY+DYVSINNLETTYNLRIHERPQIIIVANKMD+P+++ENL AFA+  
 Sbjct: 241 VILHVIDMSASEGRDPYDDYVSINNLETTYNLRIHERPQIIIVANKMDIPEAQENLAKFK 300  
 Query: 301 KLAANYDEFDMPMIFFISSLAHQGLENLADATAEILLANTEFLLYDRTDMQSEAYYGF 360  
 KLAAYDEFD+PMIFFISSLAHQGLENL++ATAELLA T+RFLLYDE+D++EAYYGF  
 Sbjct: 301 KLAAYDEFDLPMIFFISSLAHQGLENLLEATAEILLAKTDREFLLYDSLDVDEAYYGF 360  
 Query: 361 NEDERPFEITRDDDATWVLGYGDKLEGLFVMTNMRDESIMKFAQLRGMGVDEALRERGA 420  
 E B+ FEITRDDDATWVL G+KLE+LFVMTNMRDESIMKFAQLRGMGVDEALRERGA  
 Sbjct: 361 AETRKDFEITRDDDATWVLGSEKLERLFVMTNMRDESIMKFAQLRGMGVDEALRERGA 420  
 Query: 421 KDGDIVRIGNFEFEFVD 437  
 KDGD VRIG FEFEFVD  
 Sbjct: 421 KDGDVRIGKFEFEFVD 437

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1319**

A DNA sequence (GBSx1399) was identified in *S. agalactiae* <SEQ ID 4037> which encodes the amino acid sequence <SEQ ID 4038>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have a cleavable N-term signal seq.  
 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4039> which encodes the amino acid sequence <SEQ ID 4040>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have a cleavable N-term signal seq.  
 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 30/42 (71%), Positives = 37/42 (87%)  
 Query: 1 MAFGDNQQRKKTGFERKLTFLVVLIMVLVITVGLVFGAISAIM 42  
 +AFG+NG RKKT FEK+T+FVVLIMVLVITVGL+ A+S +M  
 Sbjct: 1 VAFGNGERKKTTFEKVIMFVLIMVLVITVGLIASALSVM 42

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1320**

A DNA sequence (GBSx1401) was identified in *S. agalactiae* <SEQ ID 4041> which encodes the amino acid sequence <SEQ ID 4042>. Analysis of this protein sequence reveals the following:

-1453-

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.2484 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AAD28348 GB:AF102860 aminopeptidase PepS (Streptococcus thermophilus)  
 Identities = 247/413 (59%), Positives = 313/413 (74%)
- Query: 1 MVLQDFUNLKKYQQLIISKGLNVQGHITLALIDVQVHLARLLTEAAYEKGQASEIVD 60  
 MVL +F L+KYA+L+++ G+NVQ GHT+AL+IDVQ L<sup>a</sup> L<sup>a</sup> + AY GA+EVIV
- 15 Sbjct: 1 MVLNPFKNLEKYAKLLVTNGINLVQGHITVALSIDVQAEALAHLLVKEAYALGAASEIVQ 60
- Query: 61 YTDQFITRQRLHLASDEVLLNVQYTTVDKSLALLNKASRLVVKSSNPNFATVDFKRLS 120  
 ++DD I R+R LHA + VP Y + LL KKASRL V+SS+P+AF V P+RLS
- 20 Sbjct: 61 WSDTINRERFLHAENRISEVPAYKAKMEYLLKKASRLVVRSSDDPDAFNGVAPERLS 120
- Query: 121 ETTRATAIALEQSRALQANKVSWNVAAAAGRENAALVFPFLKTSDDQVDMALMTFKLN 180  
 +A A + A Q+NKVSQ VAAAAG+ENA VFP + ++ VD L<sup>a</sup>+ IFK
- Sbjct: 121 ALAALGAGRAFKFMQVATQSNKVSQVAAAAGKQAKKVPFPAASSDEEAVLLN<sup>a</sup>QIFKTC 180
- 25 Query: 181 RYIEDOPIAAMDHEAKILKEKATRLNQEFDALHYTAPQDTL/LGMPKNHIWKAAGSINA 240  
 R+YE DP+ AW H +L KA LN+ QF ALHYTAPQDTL/LG+PKNE+WE+AG+NA
- Sbjct: 181 RYIEKDFVRANKHADRDLAKARILNEAQFSALHYTAPQDTL/LGLPKNEVWESAGASINA 240
- 30 Query: 241 QGSEFIANNPTEIEFSAFDYRRADGYVTSTKPLSYAGVLIENMTFTFKDKIINVTAKG 300  
 QGSE+F+ NMPTEB+F+AD+RRA GYV+STKPLSY G IIE + TFKDG+I+++TA+G
- Sbjct: 241 QGSEFLNMPTEIEVF<sup>a</sup>TA<sup>a</sup>DFRRAYGYVSTKPLSYNGNIIEGKIVTFTFKDGEIVDITADQG 300
- Query: 301 QETVQRILEENDGARSLGEALVPHKTPISLSGLIFNTLFDENASNHLAIGTAYAFNVE 360  
 ++ ++ L<sup>a</sup> N+GAR+LGE ALVP +PIS SG+ FNTLFDENASNHLAIG AYA +VE
- 35 Sbjct: 301 EKVMKINLVFNNGARALGECALVPDSSPISQSGITFFNTLFDENASNHLAIGAAAYATVE 360
- Query: 361 GGTMTSQQELDEAGLRSSSTHVDPMIGSEQMDIGDIDRAGTAVPIFRNGEWA 413  
 GG +MT +EL AGLNRS HVDF+IGS QM+IDGI DG+ VPIFRNG+W I
- 40 Sbjct: 361 GGADMTREELKAGALNRS<sup>a</sup>SVHVDPIIGSNQMNIDGIIHDGSRVPIFRNGDWI 413

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1321

- 45 A DNA sequence (GBSx1403) was identified in *S.agalactiae* <SEQ ID 4045> which encodes the amino acid sequence <SEQ ID 4046>. Analysis of this protein sequence reveals the following:

Possible site: 33  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -7.91 Transmembrane 661 - 677 ( 657 - 680)

----- Final Results -----

- 55 bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8787> which encodes amino acid sequence <SEQ ID 8788> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7  
 McG: Discrim Score: 6.47

-1454-

GvH: Signal Score (-7.5): 1.01  
Possible site: 29  
>>> Seems to have a cleavable N-term signal seq.  
ALOM program count: 1 value: -7.91 threshold: 0.0  
5 INTEGRAL Likelihood = -7.91 Transmembrane 658 - 673 ( 657 - 680)  
PERIPHERAL Likelihood = 4.35 555  
modified ALOM score: 2.08

\*\*\* Reasoning Step: 3

10 ----- Final Results -----  
bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15 LEXTG motif: 647-651

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:NAF09821 GB:AB001885 6-aminohexanoate-cyclic-dimer hydrolase  
[Deinococcus radiodurans]  
Identities = 150/497 (30%), Positives = 233/497 (46%), Gaps = 32/497 (6%)

Query: 110 LTESYTKKQKQDLNWMVRSQGVISEELVNMAYDIIAKENPSLNAVITTRRQEAIEBARK 169  
LT Y + D DLA + R G+++E++ A N +LNAV+ + + +AR  
25 Sbjct: 45 LTFAYYDLRLDALDLAQLPRGELSABDMCTAAIHRAQVNVNVALNAVYPLYDQSLAQARA 104

Query: 170 L-----KDTNQPFLLVPLVKGHLSIKOGEINNGLIYADGKISTFDSSTVKKYKDLG 222  
+ PF GVP LNK G + G G +I +D V+++ G  
30 Sbjct: 105 TDAARARGEQATGPFAGVPLVLDPGSLAGVPHPTGSTRAYRDQIPEWDELVRVMQAAG 164

Query: 223 FIILGQINPFPEYGRNITDSKLYGLTHNPWDLAHNAGSSGSGSAAALASGMTPIASGSDA 282  
+ LG+TN PE+ +T+ +L+G T NPWDL G3SSGSGSA+AA+G+ P+A D  
35 Sbjct: 165 LLPLGKINIPEFALMGVTEPELHGPTRNPDWDLGRTFGSSGSGSASNAAGIVPLAAGDG 224

Query: 283 GGSIRIPSSWTGLVGLKPTKGLV---SNEKDSYSTAVHFPITKSSRDAETLLTYLKSSD 339  
GGSIRIP+S GL GLKP+RG V AV LT+S RD+ LL + D  
40 Sbjct: 225 GGSIRIPASCOGLPGLKPSRGVPOGDGVGEFPAQGAIVEHVITRSVRDGAALLDLQSPD 284

Query: 340 QTLVSV-----NDLKSPLIAYTLKSPMGTESVQDARNAIMDNVTLRKQZFK 386  
+ L I ++ P+G V + A+ L G +  
45 Sbjct: 285 AGAALFLPSPERPYSEVGREGRLIGFSTAHLGRSVHPECVAAVQGAARLLESLSGE 344

Query: 387 VTEIDLPLDGRALMRDYSLAIGMGAPSTIEKDLKKHGFTKEBVDPIITWAVHYIQNSD 446  
V E+ LP DG AL + + L G GA +D DV+ +TW + LL + RA  
50 Sbjct: 345 VEEVALPMDGPALQAFLMIYGETGASIALRLTLGRPARASDVEAVTWLLQGLRSYS 404

Query: 447 KALKKGIMEAQKIMDDYRKAMEKHLKQFFIFLGSFTTASIALPLNDPY----VTEEDGKA 502  
A+ A+ + + +AM + H+ + + L+P A+ PL V RA  
55 Sbjct: 405 AAD-----FAAARASWNVHARAMGRPHONTDLLITPVIAT-PPQLQIGELQPRGVQAALLRA 459

Query: 503 IYNNENLSQERELALFNQWEPMLRKPTPTQIANTGLPAISITPYLSEGLPIGTMLMA 562  
M+ R + +L + P+TQ+AN+TG PA+S+P + + GLP+G +A  
60 Sbjct: 460 AQQMDVSGLLRRSGQDALATDILEMPYTLQANLTQSPAMSVELHWADGLFVGQVFA 519

Query: 563 GANYIMVLIKPATPFKK 579  
+ VL++ A E+  
65 Sbjct: 520 PLAREDWLRLAQLGQSQ 536

There is also homology to SEQ ID 4048.

- 60 SEQ ID 8788 (GBS173) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 5; MW 96.8kDa).

The GBS173-GST fusion product was purified (Figure 116A; see also Figure 201, lane 7) and used to immunise mice (lane 1+2 product; 15µg/mouse). The resulting antiserum was used for Western blot, FACS,

-1455-

and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 1322

A DNA sequence (GBSx1404) was identified in *S. agalactiae* <SEQ ID 4049> which encodes the amino acid sequence <SEQ ID 4050>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (rsuA). Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3674 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA006992 GB:AF001518 168 pseudouridylate synthase [Bacillus halodurans]  
 Identities = 110/236 (46%), Positives = 149/236 (62%), Gaps = 4/236 (1%)  
 20 Query: 1 MRIDKFLVBCGLGSRTQVKLILKKQISVNGNSETSPKVQVDEYDEIKYNGTLVSYEKF 60  
 WR+DKFL G GSR VK +LK + V G P V+ + I G V Y+ +  
 Sbjct: 1 MRIDKFLANMGFSRKDKVKLLKRTGAVRVCQGPDKPSTHVFEPRESITVYGEVEYKPY 60  
 25 Query: 61 VYMLHKKPKGVISATDDPSKTYLDDLLDKTARDKAVFPVGRLLDITDGLLLTNNGLAH 120  
 VY M++KPKGVI AT+D H+TV+DLL + R PVGRLL DT GLL+TN+G+ H  
 Sbjct: 61 VYLMNKKPKGVICATEDLEHETVIDLGGEEERHYEYSPVGRLLKDTVGLLLITNDGKPNH 120  
 30 Query: 121 KMLSPKGVHDKCYEVKISQIMTEDDILAFDKGIILKD-PTCLFALLEIVEVNVQVKQSLV 179  
 ++SFK HV K Y + G +TE+D+ AE G++L D + EA L I+ + +S +  
 Sbjct: 121 WLMSEKGVHFKYTRALVBGHVTEEDVGAFSHGVVLDGQVTKATLHLEAG---ARSH 177  
 Query: 180 KITIKRQKPHQVKRMVAACGKEVLEKRLRMGLQDKQLESQWRRLTIKEIEKL 235  
 ++ + EGKPHQVKRM A GK VLEL+R+++GML LD +L G++R LT +EI L  
 35 Sbjct: 178 ELILTEGKPHQVKRMFQAVGKRVLELERIKIGMLLDPELARGYERLATEALL 233

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4051> which encodes the amino acid sequence <SEQ ID 4052>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0152 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA090821 GB:AE001895 6-aminohexanoate-cyclic-dimer hydrolase  
 [Deinococcus radiodurans]  
 50 Identities = 177/485 (36%), Positives = 259/485 (52%), Gaps = 13/485 (2%)  
 Query: 5 DATAMAIIVATGQTPLPVLVQAIYKAKKNPTLNAITSERFEALAEAKORDFSGL--- 61  
 DA +A + G+ + ++ T AI++A+ N LNA+ ++ L +A+ D +  
 Sbjct: 54 DALDLALQFRGELSEARIMCTAAIHRAQVNVNALNAVVPFLYDQGLAQARATDAARARGE 113  
 55 Query: 62 ---PFAGVPLFADKLDQQLKQHSSTSGSLFKKEYQATKTLDFVKRLALGFIILGRSNT 117  
 PFAGVP +KD G L G T G+R +++ D V+R +A G + LG++NT

-1456-

Sbjct: 114 QATGPFAGVFLVKDFGSRLAGVPHGTGGTRAYRDQIPENDDLVRRWQAGLLPLQKINT 173

Query: 118 PEPGFKNISDSSLKGPVNLPRDNTNRAGGSSGGAALVSSGISALATASDQGGSTIRIPAS 177  
PEF +++ LKGP F D R GGSSGG+A+ V++GI LA A DGGSGSTRIPAS

5 Sbjct: 174 PEPALMGVTEPELHGPTNRPMDLGRITPGSSSGGSASAVAAGIVPLAGHGQGGSTIRIPAS 233

Query: 178 FRNLGLGLKPSRGRMFVPGSGYSRWQASVHFALTSVRDTRNLLYLQMGBMESPPFLAT 237  
GL GLKPSIR+P G G WQGA+V IT+SVRD+ LL Q + L +

10 Sbjct: 234 CGSLGLGLKPSIRKVPQSDGVGEWQGA+VHVHLTRSVRDSAAALLDLBGGPDAGALLPLAS 293

Query: 238 LTKDSIYQSLQRP--LITAFYQRLSDGSPVSLDTAKALRQAVTWLRBQGHQVLEEEFFV 295  
+ + + P L I F G V + A++A L GH++E+ P

Sbjct: 294 PERFVSEEVGREGRFLRIGFSTAHLGRSVIIEPCVAAVQGAARLLSLGHEVEEV-ALFW 352

15 Query: 296 NMTEVIRHYIYINSVETAMFADIEDTFGRPMTKDDMETWAIYQSGKDIKPNRYSCVI 355  
+ + + + + ET A A + DT GRP D+E +TW + Q G+ A ++

Sbjct: 353 DGPALAQAFIMLYPGETGASLAALRDLGRPARASDVAVTWLLGQLGRSYSAADFAAR 412

Query: 356 QKWDYTSATMASFHETYDLLLTPTNTNPAKKGELVP---DSKLMANLAQZIPSSEHQF 412  
N+ ++ M FH+ YDLLLT TP + GEL F + L+ Q ++ +

20 Sbjct: 413 ASNVHARAMGRPHQNTDLLLTFLVLTATPLQIGELQPRGVQALLRAAQMDVSGLLRS 472

Query: 413 NLVETMPGKSLAINPTALPNLTQQAISLPTYETKGLSGHGIQLIAKGRDRLLLGIAB 472  
V+ + L PTT L NLRQQA+S+P + T +GL +G+Q +A RED+LL +A

25 Sbjct: 473 QVQDALATDILEKMPYTLANLTQQPAMSVPFWHTADGLFVGVQFVAPLAREDVLLRLAG 532

Query: 473 QFZAA 477  
Q E A

Sbjct: 533 QLEQA 537

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/240 (62%), Positives = 183/240 (75%)

Query: 1 MRLDKFLVBOGLGSRTOVKLLIKKKQISVNGNSETSPKVQVDVEYDEIKYNTIVSYK 60  
Sbjct: 2 MRLDKFLV G+G+R+QVK+LKKK I VN ETS K +DEX+D + Y GT + YE F  
MRLDKFLVATGVQTRSGVKLLIKKAI+VNVKQVETSAKAHIDEYKDLVITYGTPVLYESS 61

Query: 61 VYMLHKPKGVISATDDPSHKTVLDLDDKTARDKAVFPVGRLEDITDTGLLLLTNNGLAH 120  
VYY+L+KP G +SAT D TV++LLD TAR KAVFPVGRLED DT GLLLLTNNG+LAH

40 Sbjct: 62 VYYLNLKPSGYSATQDRQQATVMEILLDDTARQKAVFPVGRLEDKTRGLLLLTNNGLAH 121

Query: 121 KMSPKGVYDKCYEVIKSGIMTEDDILAFDKGIIILKDFCTCLPALLEIVEVNVKQKSLVK 180  
+LSPKHHV K Y K++GIMTE D F +GI LKD CLPA LE++ + ++ SLVK

Sbjct: 122 DLLSPKHHVKEYLAKVAGIMTEADKDYFARGISLKHQCLPAHLEVLASDLQKQSLVK 181

45 Query: 181 ITIKBKPHQVQKRMVAACGKEVLKRLRLMNLQDKQLBSCQWRRITIKETIKLEKYNQ 240  
ITI+BGKPHQVQKRMVAACGKEVL+L+RL MG L+LD L G++RRLT +E++ L Y Q

Sbjct: 182 ITIQBKPHQVQKRMVAACGKEVLQRLSMGPLKLDPSLABGEFRLTPEELQSLAPYQ 241

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1323

A DNA sequence (GBSx1405) was identified in *Sagalactiae* <SEQ ID 4053> which encodes the amino acid sequence <SEQ ID 4054>. Analysis of this protein sequence reveals the following:

55 Possible site: 46  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.2811 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1457-

A related GBS nucleic acid sequence <SEQ ID 10007> which encodes amino acid sequence <SEQ ID 10008> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AA57350 GB:J04483 reductase [Leishmania major]  
Identities = 129/277 (46%), Positives = 167/277 (59%), Gaps = 3/277 (1%)

Query: 26 TLENTLNIPIKIGPGTQWLTGSEAYKAVTHALKGVYTHIDTAQIYQNEHSVGRAIRDSGL 85  
TLEN + +P+ G G WQ GR AV AL CY HIDTA IY NE SVG +R SG+  
Sbjct: 10 TLENGVIMPGQLGVLGWSGSPACSEVTNNAVHWALCAGYRHIDTAATYKNEBSVGAIGLRASG 69

10 Query: 86 ARRESIFLITIKWINDKHIDHLAKASIDESLQKLGVDYIDLLIHWNPKALRENDAMKAGN 145  
RE +F+ITK+WN + Y A+ +ES QKLGVDYIDL LIHW K + + K  
Sbjct: 70 FREDVFITIKLWNTQGVVESTLAAPKESRQKLGVDYIDLYLIHWMPGKIDLSKEGKY-- 127

15 Query: 146 AGTWKAMEBAYKEGKVKAIQVSNFMKHLEALFETAEIKPMVWNIILAPGCAQEDLWRF 205  
+W+A E+ YKE KY+AIQVSNF HHLE + + PMVWQ+ L P Q DL FC  
Sbjct: 128 LDSWRAPEQLYKEKVKRAIGVSNFHHLELDVLAACITVPMVWQVELHPIALQADLRAPC 187

20 Query: 206 KGNIDILEAYSFGPTGAIFENHSIKAIARCKYKSVACVALRWSLDMGFLPLPKSATPKNY 265  
I +ED+SP G G + N + AI KY K+ AQV LRW++ + +PKS + I  
Sbjct: 188 DAQIKYEAMSPQAQKLLSNFILSAIGAKYKNTAQAQVILRWNIQKLTITPKSVHRI 247

Query: 266 EANDLIDFDFQNEEDIALTLQLDSGIK-PKDPANVSF 301  
E N DIFDF+L +D+ ++ L++ + DFD F  
Sbjct: 248 EENADIDFDFELGAEDVMSIDALNTNSRYGDFDFDEAQF 284

25

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 779> which encodes the amino acid sequence <SEQ ID 780>. Analysis of this protein sequence reveals the following:

Possible site: 27  
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0980(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/282 (54%), Positives = 204/282 (71%), Gaps = 2/282 (0%)

40 Query: 20 IVMTETLNTLNIPIKIGPGTQWLTGSEAYKAVTHALKGVYTHIDTAQIYQNEHSVGR 79  
+++ T +++ IP +GPGT+Q +GEEAY+ A+K GY HIDTA IY NE SVGRA  
Sbjct: 1 VMVTTVKGTSGYIPIVLPFGTYQADGSEAYQSTLAALIKAGYRHIDTAATYKNEBSVGR 60

45 Query: 80 IRDGLARESIPIATIKWINDKHIDHLAKASIDESLQKLGVDYIDLLIHWNPKALREND 139  
I+DSG+ RE +F+ITK+WND E Y AK ++ SL +LG+DY+DL LIHWNPALR +  
Sbjct: 61 IKDSGLVRKDLFITTKLWANDHSTYHAKDALAASLRLGLGVDYIDLYLIHWNPALR--N 118

Query: 140 AWKAGNAGTWKAMEBAYKEGKVKAIQVSNFMKHLEALFETAEIKPMVWNIILAPGCAQ 199  
WK NA W+ MESA + G +AIQVSNFM HLEAL ETA+I P +NQI LAPGC Q+  
Sbjct: 119 TWKEANAQAMQMEBAYEAGLIKSGVSNFMVHLEALQETAKITPAINQIRLAPCYQR 178

50 Query: 200 DLVRPCKNGIDILEAYSFGPTGAIFENHSIKAIARCKYKSVACVALRWSLDMGFLPLPKS 259  
+V+ +CK N+ILLR+SP G G IP+NE+++ +A KY K+VAQVAL WSL GF+PLEKS  
Sbjct: 179 EVDVYCKANILLR+SPGQEIFPDNEMQQLANKYKDTVAQVALMSLAPGIFLPLKS 238

55 Query: 260 ATPKTEANLIDFDFQNEEDIALTLQLDSGIKPKDPANVSF 301  
+ I+ N+ IFD L +D T L + +PD SF  
Sbjct: 239 VHDRIKENMALFDVSLTQEDKTIKYLSGMSALINFDITSE 280

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1458-

**Example 1324**

A DNA sequence (GBSx1406) was identified in *S. agalactiae* <SEQ ID 4055> which encodes the amino acid sequence <SEQ ID 4056>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0633 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10009> which encodes amino acid sequence <SEQ ID 10010> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:CAB12612 GB:Z99108 similar to NAD(P)H-flavin oxidoreductase
 [Bacillus subtilis]
 Identities = 106/223 (47%), Positives = 150/223 (66%), Gaps = 8/223 (3%)

20 Query: 29 DIKKQVRRAFDPRMAIRVYN--NDIPKEDMEYILDTAMLSPPSSVGLGWRFLVLDNQITIA 87
 D+K Q+ A++FR A++ N + D E+IL+T LSPSS+GLE W+F+V+
 Sbjct: 3 DLTKQLDQYNFRHATKGFDPKPKVSDSDFEFLTETGRLSPPSSIGLEGWKFVVDNQNP--- 59

 Query: 88 KFRDKLKEVANGAQYQOLDTASHFVLLAE--KGATYNADSMINSILIRGLGDPAALESRI 145
 +FR+KL+E WQAQ QL TASHFVL+LA K YNAD + L E+
 Sbjct: 60 EFRKELREYTWGAQQLFTASHFVLLARTAKDIKNADYIKRHLKEVKQMPPQVVEGYL 119

 Query: 146 PLYKSFQENDKI--DSERSLMDWTAKQTYIALGNMNTAAAMIGVDSCPIEGFDYEKVNIN 204
 + FQ+ND+ + S+R+L+DW +KQTYIALGNMNTAAAMIGVDSCPIEGF Y+ + I
 Sbjct: 120 SKTESFQKNDLHLESDRTPDWAKSQTYIALGNMNTAAAMIGVDSCPIEGFQYDHIHRI 179

30 Query: 205 LSKGLLDDKKEALSCMVSPGYRLREPKHSRARKERQEVITWV 247
 L+EG+L++ IS MV+FGYR+R+D+ + R ++V+ WV
 Sbjct: 180 LEEEGLENGSFDISVMVAFGYRVDPFR-FKTRSAVEDVVDV 221

```

35 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4057> which encodes the amino acid sequence <SEQ ID 4058>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1705 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 An alignment of the GAS and GBS proteins is shown below.

```

Identities = 126/222 (56%), Positives = 174/222 (77%), Gaps = 4/222 (1%)

Query: 28 EDIKQVRRAFDPRMAIRVYNNDIPKEDMEYILDTAMLSPPSSVGLGWRFLVLDNQITIA 87
+ I Q++A FR A+RVY I ED+ ILD ANLSPSS+GLEGWRFLVLD + I
50 Sbjct: 3 QTHHQIQALHFRTAARVYVYKEEKISDEDLAILDAWLSPPSIGLEGWKFVVDNQNP--- 61

 Query: 88 KFRDKLKEVANGAQYQOLDTASHFVLLAEKGYYNADSMINSILIRGLGDPAALESRIPL 147
 +++K K ANGAQYQL+TASHF+LLAEK A Y++ ++ NSL+ENG+ + L SR+ L
 Sbjct: 62 --KEKIKPFAWGAQYQOLDTASHFILLIARKHARYDSPAIKNSLLRGITKEGDLNSRLK 119

55 Query: 148 YKSFQENDKI--DSERSLMDWTAKQTYIALGNMNTAAAMIGVDSCPIEGFDYEKVNINLS 206
Y+SFQ+ DN + D+ R+L+DWTAKQTYIALGNMNTAAAMIGVDSCPIEGF Y+KVN+IL+
Sbjct: 120 YESFQKEDMDALNPRALFDWTAKQTYIALGNMNTAAALGIDTCTPIEGFHYDKVNHILA 179

```



-1459-

Query: 207 REGLIDDKKEAISQVSPGYRLREPKHRSRKRQEVITWVE 248  
 K +ID +KE I+ M+S GYRLR+FKH++ RK ++EVI+ V+  
 Sbjct: 180 KIRVIDLEKSGTASMLSLGYRLRDPQIAQVRKPKKEIVSVVK 221

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1325

- A DNA sequence (GBSx1407) was identified in *S.agalactiae* <SEQ ID 4059> which encodes the amino acid sequence <SEQ ID 4060>. This protein is predicted to be lactoylglutathione lyase (gloA). Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1656 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP.AAC21986 GB:U32717 lactoylglutathione lyase (gloA) [Haemophilus influenzae Rd]  
 Identities = 59/131 (45%), Positives = 86/131 (65%), Gaps = 2/131 (1%)  
 Query: 1 MPFLHTCIRVKDLDSIAFYQALGFKEVRRNDFFENQPTLVYMALEDPSY-ELHLYTN 59  
 M LHT +RV DLD SI FYQ+ LG ++R ++ PE ++TL ++ ED S E-ELTYN  
 25 Sbjct: 1 NQILHTWLRVGLDLSIKFYQDVLGMRLLRTSENSEYKYTLFLAPLEGDSAAEILTYN 60  
 Query: 60 YDHEAYDLNGYGHIAVGVDLETTYDAHQKAGYSVFKISG-LPGKPMFYFIQDPDGYK 118  
 + + X+ G YGHIA+GVDD+ T +A + +G +VT+ +G + G + F++DPDGYK  
 30 Sbjct: 61 HGVDKYEHGTAYGHIAIGVDDIYATCEAVRASGQNVTRAGPVKGGSTVIAFVEDPDGYK 120  
 Query: 119 IEVIRLSQFKA 129  
 IE I K+  
 Sbjct: 121 IEPIENKSTKS 131

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4061> which encodes the amino acid sequence <SEQ ID 4062>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1382 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 80/125 (64%), Positives = 93/125 (74%), Gaps = 1/125 (0%)  
 Query: 1 MPFLHTCIRVKDLDSIAFYQALGFKEVRRNDFFENQPTLVYMALEDPSY-ELHLYTN 60  
 M LHTCIRVKDLDSIAFY A FKE R DFP++QPTLVY+ALE +SYELELTNY  
 50 Sbjct: 1 MKALHTCIRVKDLDSVAFPTSAPFPKENVYKDFPDQPTLVYALAGE-SYELELTNY 59  
 Query: 61 DHEAYDLNGYGHIAVGVDLETTYDAHQKAGYSVFKISGLPGKPMFYFIQDPDGYKIE 120  
 H YDLNGYGHIA+G + R + H++AG+ VT I L K +YFIQDPDGYKIE  
 Sbjct: 60 GHGVDLNGYGHIALGSEHFEADHKHQRQGFPPVTDIKELADKSARYFYFIQDPDGYKIE 119  
 55 Query: 121 VIRLS 125  
 VI L+  
 Sbjct: 120 VIDIAN 124

-1460-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1326

- 5 A DNA sequence (GBSx1408) was identified in *S.agalactiae* <SEQ ID 4063> which encodes the amino acid sequence <SEQ ID 4064>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
10 INTEGRAL Likelihood = -9.02 Transmembrane 241 - 257 (229 - 262)
 INTEGRAL Likelihood = -4.94 Transmembrane 270 - 286 (264 - 287)

----- Final Results -----
 bacterial membrane --- Certainty=0.4609 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB12688 GB:Z99108 stress response protein [Bacillus subtilis]
Identities = 139/304 (45%), Positives = 200/304 (65%), Gaps = 3/304 (0%)

20 Query: 3 LLSVIVPCYNEQETVSTFLTEIKKVESEMARYTHFEYIVNDGSTDRTELLKKAQKQPD 62
 L+S+I+P YNE V +KK E+ Y +E F+NDGS D TL+ +K A
 Sbjct: 5 LISIIIPSYNEGYNV/KLHESLRK-KFQNIHYD-YEIFFINDGSVDUTLQQIKDLAATCS 62

25 Query: 63 NVHYLSPSRHFORDAALLAGLSHTTGDFFITVMDVQLQDPPTLLPEMYLKLQSGYDIVATR 122
 V Y+SPSR+FGK+AA+LAG EH G+ + VMD DLQ P LL E +EGYD V +
 Sbjct: 63 RVKYSISPSRNFGRKAAILAGFHFVQGEAVTMDAILQHFYTLKKEFIKGYREGYDQVIA 122

30 Query: 123 RKDRKGEPLIRSLFAKLFYKLIQVSDTQVVDGARDFLMTRKQVDSILELNEVNRFSKG 182
 R +RKG+ +RSL + ++YK IN+ + + DG DFL+++Q V+++L+L+E NRFSKG
 Sbjct: 123 R-NRKGDSFVRSLLSGMYKFINQAVEVDLRDGVGDFLLSRQAVNALLKLEGNRFSKG 181

35 Query: 183 IFSWIGYDVAYISYENRERIAGKTSWSFFNLLKYSLDGFINFSEIPLAIATWIGTLLSVL 242
 +F WIG+D +YEN ER G +WSP +L Y +DG ++F+ FL + +G +L
 Sbjct: 182 LFCWIGFDQKIVFYENVERKNGTSKNSPSSLFNYGMDGVSPNHKPLRLCFTTGIFILL 241

40 Query: 243 SLALIIIFIIRKLLFGDPVSGWASTVTIVLFWGIGIQLLSLGIIGIKYISKIFLETKKRPVY 302
 S++ II ++ L G V G+ ++ VL F+GG+QLLSLGIIG+YI +I+ ETYKRP Y
 Sbjct: 242 STIYIATFYKILTWGISVPGYPTIISAVLFLGGVQLLSLGIIGIYGRITYETKGRPHY 301

45 Query: 303 IVKE 306
 ++KE
 Sbjct: 302 LIKE 305

```

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4065> which encodes the amino acid sequence <SEQ ID 4066>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
50 INTEGRAL Likelihood = -9.55 Transmembrane 256 - 272 (251 - 282)
 INTEGRAL Likelihood = -5.31 Transmembrane 290 - 306 (284 - 307)

----- Final Results -----
 bacterial membrane --- Certainty=0.4821 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
55 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9113> which encodes the amino acid sequence <SEQ ID 9114>. Analysis of this protein sequence reveals the following:

-1461-

Possible cleavage site: 36  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty= 0.482 (Affirmative) < succ>  
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 207/307 (67%), Positives = 258/307 (83%)

Query: 1 MALLSVIVPCYNEQETVSTFLTEIKKVSSEMARYTHFEYIFVNDGSDTRTLELLKGAQK 60  
 M LLS+IVPC+NE+ + + E+ ++E+ M FEYIF+DGS D TL +L++ A +  
 Sbjct: 21 MTLISIIVPCNEENILPYFERMHQLETSMTNQLAPEYIFIDGSKDNTLIGILRELAAR 80

Query: 61 FDNVHYSFSRHFQKQALLAGLEHTTGDFITVMDVDLQDPPTLLPEMYLKLGQSYDIVA 120  
 F NVHYSFSRHFQK+ A LLAGL+ G++ITVMDVDLQDP LLP MY KL+SGYDIV  
 Sbjct: 81 FNVHYSFSRHFQKAGLLAGLKEAGNYITVMDVDLQDPPELLPIMYAKLKEGYDIVG 140

Query: 121 TRKDRKGEPLIRSLFAKLFYKLINQSDTRKVDGARDFRMTKQVVDSEILENEVNRFS 180  
 TRR++R+GEPLIR+ + LFY LI +SDT+MV+G RD+RLMT+QVVDSEILE EVNRFS  
 Sbjct: 141 TRQNQRGEPLIRSMCSNLFYGLIKHLSUTRMVNGVRDYRI NTRQVVDSEILEGEVNRFS 200

Query: 181 KGIFSNIGYDVAYISYENRERIAGTKTSSFFNLLKYSLDGFINFSEIPLAIATWIGTLSS 240  
 KGIFSN+GY + Y+S+EN++R GK+ W F+ LL+YSLDGFINSF+PL IATW GT S  
 Sbjct: 201 KGIFSNVGYRITLYSPFNQKRYKYGSRMHFWELLKYSLDGFINFSEIPLATWGTGTF 260

Query: 241 VLSLAIIFIIRKLLFGDPVSGNASTVTIIVLFWGGIQLLSLGIIGKYISKIFLETKR 300  
 ++S+ AL+FIIRK+LFGDPVSGNASTV+I+LFWGGIQL +GIIGKYISKIFLETKR  
 Sbjct: 261 LISIFAILFIIRKILKLLFGDPVSGNASTVSIILFWGGIQLFCMGIIGKYISKIFLETKR 320

Query: 301 VYIVKEE 307  
 +YI+KE+  
 Sbjct: 321 LYIIEKE 327

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1327

A DNA sequence (GBSx1409) was identified in *S. agalactiae* <SEQ ID 4067> which encodes the amino acid sequence <SEQ ID 4068>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.44 Transmembrane 50 - 66 ( 50 - 66)  
 INTEGRAL Likelihood = -1.49 Transmembrane 27 - 43 ( 27 - 43)

----- Final Results -----

bacterial membrane --- Certainty=0.1977 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA83253 GB:Z31377 potential amino acid permease  
 [Lactobacillus delbrueckii]

Identities = 34/55 (61%), Positives = 44/55 (79%)

Query: 7 DHTQSENGNVRGLNREHVQLIAIAGTIGTGLPLGAGRSISLTGSPSIVLVAITG 61  
 D + ++ +G +R L NRHVQ+IAI GTIGTGLPLGAS +IS TGP8++ +YAI G  
 Sbjct: 5 DRSIENTDGTIRLSLRNREVQMIAGSTIGTGLPLGAGTTISATGSPSIVTIAIMG 59

-1462-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4069> which encodes the amino acid sequence <SEQ ID 4070>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence
5 INTEGRAL Likelihood = -11.15 Transmembrane 170 - 186 (161 - 190)
 INTEGRAL Likelihood = -8.44 Transmembrane 256 - 272 (252 - 274)
 INTEGRAL Likelihood = -8.33 Transmembrane 352 - 368 (347 - 375)
 INTEGRAL Likelihood = -7.54 Transmembrane 139 - 155 (133 - 160)
10 INTEGRAL Likelihood = -5.73 Transmembrane 420 - 436 (417 - 440)
 INTEGRAL Likelihood = -3.88 Transmembrane 56 - 72 (54 - 75)
 INTEGRAL Likelihood = -3.40 Transmembrane 283 - 299 (282 - 300)
 INTEGRAL Likelihood = -3.29 Transmembrane 440 - 456 (439 - 458)
 INTEGRAL Likelihood = -1.49 Transmembrane 31 - 47 (31 - 47)
15 INTEGRAL Likelihood = -1.33 Transmembrane 109 - 125 (109 - 127)

----- Final Results -----
 bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAB14651 GB:Z991:7 amino acid permease [Bacillus subtilis]
Identities = 210/454 (46%), Positives = 296/454 (64%), Gaps = 11/454 (2%)

25 Query: 12 DNNELNGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIFVVMITGAPMFM 71
 DN + + RGL+NRH+QL+AI G IGTGLFLG+G+SI GPSI+P Y+ITG F P +
 Sbjct: 8 DNFQQQLKSRGLKNRHQLMAIGGAIGTGLFLGSGKSHFAGPSILFAYLITGVCFPT 67

30 Query: 72 MRAIGEMLYDPDQHTFIFNISKYIGPGWGYFSGLSWISLFIQARITAVGAYVQFWF 131
 +R+GE+L + H+P++F+ Y+G + +G +YW I + NA++TAVG Y Q+W
 Sbjct: 68 IRLSGELLSSAGHSGYFDFVDYDLGNMAAFITGWTNFWCHSLMADLTAVGIYTCQWL 127

35 Query: 132 PSNPAMLIQVLVPLVLLSSINLIARVFGSTEFWPMIKILAILALIAITAPMLTGPETH 191
 P P WL L+ L++L +NL V++GGE EFWEA+IK++AILALI T I ++ GF
 Sbjct: 128 FDFVQLVPLGLALITILLMLLATVKLGELEFWALIKVIALIALIVTGILLIKGPSAA 187

40 Query: 192 TGHASLGNIFDHPSPFNGKLPFMAFQMVFFAYQALEFVGITTSSETANPRKVLDAIQE 251
 +G ASL+N++ H MFPNG F++QGV FA+ IE VG+T ET NP+KV+PKAI +
 Sbjct: 188 SGPASLNNLMSHGGMPPNGHGFILSPQMVFVAFVGLIELVGLTAGRTENPQKVPAKINQ 247

45 Query: 252 IPTRIVIFVQALVSIMATVPMHQLPVDESPPVNVFKLIGKWAALINPVLTSAASAL 311
 IP RI++FYVQAL IM I PW+ L +ESPFP VP +GI AA+LINPVLTSAASA
 Sbjct: 248 IPRVILLFVQALFVIMCTYPMVNLMPNESPFVQVPSAVGVVAASLINPVLTSAASA 307

50 Query: 312 NSTLYSTGRHLYQIANE--TFNALNRLKINTLSRQGVPSRAIIASAVVIGSALINILP 369
 NS L+ST R+Y +A+ P L L+ VPS A+ S++ I +N L
 Sbjct: 308 NSALPSTSRMVYSIANDHHAQGLL-----KKLTSSIVPSNALPSSIALIGVSLNYIM 361

 Query: 370 GVADAFSLTASSGGVYIAIYALIMAHKWRQSK--DPMADGYLWPKYKVTPLPLAPP 427
 P+LIT+ S+ +I I+ +T+I H KYR+++ + A+ +MF Y+ + LITAF
 Sbjct: 362 -PEQVPTLITSVSTICFPIFGITVICHLYKRTKQHEAKANKFMNPFPLSNLILFL 420

 Query: 428 AFVVISLPLQSTYIGACIGATIMIIIPGIYSNVK 461
 AF+ + L L L T I +W ++ I V+
 Sbjct: 421 AFILVIALANDTRIALFVFPVFWVLLIILYKVQ 454

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 48/62 (77%), Positives = 51/62 (81%)

60 Query: 1 MSKNNNDHTQKSENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIVVIAITGA 62
 MS + ENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSI+ITGPSI+ VY ITGA
 Sbjct: 5 MSIKQTDNNELNGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIFVVMITGA 66

```

-1463-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1328

A DNA sequence (GBSx1411) was identified in *S. agalactiae* <SEQ ID 4071> which encodes the amino acid sequence <SEQ ID 4072>. This protein is predicted to be alkylphosphonate uptake protein (phnA).

Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0965 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77069 GB:AE000483 orf, hypothetical protein [Escherichia coli K12]
Identities = 79/110 (71%), Positives = 91/110 (81%), Gaps = 1/110 (0%)

Query: 1 MSLPNCPCKNSFYVEDGILLVCPFCAYENFEE-IEEEVGLVLDNGTRLSDGDTTVT 59
 MSLP+CPKCNSEY YED +CFCAYENW E +E LIV D+NG L+DGD+VT+
Sbjct: 1 MSLPHCFCKNSEYTYEDNGMYICFCAYENDAEPQAQESDELIVKDGANLLADGDSVTI 60

Query: 60 IKDLKVKGAPKDIKQSTRVKNIRLVGDHNIDCKIDGFGAMKLESEFVKK 109
 IKDLKVKG+ +K GT+VKNIRLV+GDHNIDCKIDGFG MKLKSEFVKK
Sbjct: 61 IKDLKVKGSGSMLKIGTKVKNIRLVGDHNIDCKIDGFGPMKLESEFVKK 110
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4073> which encodes the amino acid sequence <SEQ ID 4074>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3428 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 73/85 (85%), Positives = 79/85 (92%), Gaps = 1/85 (1%)

Query: 26 CAYENP-EEIESRWGLVLDNGTRLSDGDTTVTVIKDLKVKGAPKDIKQSTRVKNIRLV 84
 CA+EW P EE EE GL+VLDNG RLSDGDT+TV+KDLKVKGAPKDIKQSTRVKNIRLV
Sbjct: 2 CAFETVPGREATTEEGSLVLDNGVRLSDGDTITTVVKDLKVKGAPKDLKQSTRVKNIRLV 61

Query: 85 DGDHNIDCKIDGFGAMKLESEFVKK 109
 +GDHNIDCKIDGFGAMKLESEFVKK
Sbjct: 62 EGDHNIDCKIDGFGAMKLESEFVKK 86
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1329

A DNA sequence (GBSx1412) was identified in *S. agalactiae* <SEQ ID 4075> which encodes the amino acid sequence <SEQ ID 4076>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
```

-1464-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 1330

A DNA sequence (GBSx1414) was identified in *S.agalactiae* <SEQ ID 4077> which encodes the amino acid sequence <SEQ ID 4078>. Analysis of this protein sequence reveals the following:

15       Possible site: 13

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

          Integral   Likelihood = -0.11   Transmembrane   558 - 574 ( 558 - 574)

----- Final Results -----

20           bacterial membrane --- Certainty=0.1044(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25       >GP: CAB11971 GB: Z99105 L-glutamine-D-fructose-6-phosphate  
           amidotransferase [Bacillus subtilis]  
       Identities = 355/604 (58%), Positives = 445/604 (72%), Gaps = 4/604 (0%)

30       Query: 1   MOGI VGVGNTNATDILIQGLEKLEYRGYDSAGIFVVGDNKSQLVKSVGRIAFQKAVGD 60  
           MOGI V G +A +IL+GLEKLEYRGYDSAGI V + + K GRI A+++ V  
       Sbjct: 1   MOGI VGYIGQLDAKEILLKGLKLEYRGYDSAGIAVANEQGIHVFERKGRADIAREVVDA 60

35       Query: 61   SVSGITGIGHTRWATHGKPTBGMNAHPHTSGSGRFVLVHNGVIENYLQIKETYLTKGNLKG 120  
           +V   GIGITRWATHG+P+ NAHPH S GRF LVHNGVIENY+Q+K+ YL LK  
       Sbjct: 61   NVEAKAGIGHTRWATHGEPSTYMAHPHQAALGRFPTLVHNGVIENYVQLKQETLQDVELS 120

40       Query: 121   ETDTETAIHLVEHFVEEENLSVLEAPFKALNIIBGSYAPALIDSQDADTIYVAKNKSPLL 180  
           +TDT+ + ++E FV L EAP+K L +++GSTA AL D+ + +TI+VAKNKSPLL  
       Sbjct: 121   DITDTSEVVVQIQFVN-GLGETEAPFKRTLLKGSYALALFQNDKRETI FVAKNKSPLL 179

45       Query: 181   IGLGNGYNMVCSDMMAMIRETSEYHSHDKELVIVKKDSVEVDQYDGNVIERGSGYTAELD 240  
           +GLG+ +N+V SDAMFM++ T+Y+E+ DKE+VIV D V +++ DG+VI R SY AELD  
       Sbjct: 180   VGLGTFENNVASDAMAMLVQVMEYVLEMDKEMVITDDQVTKNLGDGVITRASYLELD 239

50       Query: 241   LSDIGKGTYPFFYMLKEIDQPTVMKRLISTYANESQDMMVDSDIKSVOEADRLYLILAAG 300  
           SDI KGTYP YMLKE DEQP VMRK+I TY +R+G+ +V DI +V EADR+YI+ G  
       Sbjct: 240   ASDIKGTYPHYMLKETDEQPVVMKRLIQTYQDENGKLSVPGDIAAFAADRLYILVQ 299

55       Query: 301   TSYHAGPAAKTMIKLIITPVFELGVSSBNGYNMPLLSKKPMFLLSQSGETADSRVLVVK 360  
           TSYHAG K IE + FVE+ V+SE+ YNMPLLSKKP+FI LQSGSETADSR VLV+  
       Sbjct: 300   TSYHAGLVGRQYIEWNVVPEVHVASFESYNMPLLSKKPLFI LQSGSETADSRVAVLQ 359

60       Query: 361   ANEMGIPSLTITNVPGSTLSREAYITMLIHAGPEIAVASTKAYTAQVATLAPLAKAVGEA 420  
           +G +LITITNVPGSTLSREA YI+L+HAGPEIAVASTKAYTAQ+LA LA +  
       Sbjct: 360   VIALGHKALITITNVPGSTLSREADYITLLHAGPEIAVASTKAYTAQVATLAVLAVSADK 419

65       Query: 421   NGGAERQDFDLVHLSIVAQSIEATLSKDVISEKVBQLLITRNAPFYIGRGNFYVTME 480  
           NS   FDLV ED I A ++EA +KD + + L +RNAF+IGRG DY+ +E  
       Sbjct: 420   NGINIG--FDLVKELGLAANAMEALCTQKDEMRMIAREYLTVSRNAFYIGRGLDYFVCGE 477

60

-1465-

Query: 481 AALKKEISYIQEGFANGELKHGTISLIEQTFVIALISADSTIAHGTGNIQEVVSRG 540  
 ALKKEISYIQ EGFA GELKHGTI+LIE TPV AL + + + RGN++EV +RG  
 Sbjct: 478 GALKKEISYIQEGFANGELKHGTIALIEQTFVIALATQEH-VNLIRGNVKEVAARG 536

5 Query: 541 ANALIIVESELERGGDDIIVNKHVPFLSAISWVITPQLIAYTASLQGLVDVKPRNLAKA 600  
 AN II +GL+ .D ++ +V+P L+ + V+P QLIAYTA+L RG DVDKPRNLAK+  
 Sbjct: 537 ANTCTISLEGLDADDRFVLPEVNFALPLVSVFLQLIAYTAALHRSCEVDKPRNLAKS 596

10 Query: 601 VTVE 604  
 VTVE  
 Sbjct: 597 VTVE 600

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4079> which encodes the amino acid sequence <SEQ ID 4080>. Analysis of this protein sequence reveals the following:

15 Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.06 Transmembrane 558 - 574 ( 558 - 574)

20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:CAE11971 GB:Z99105 L-glutamine-D-fructose-6-phosphate  
 amidotransferase [Bacillus subtilis]  
 Identities = 353/604 (58%), Positives = 445/604 (73%), Gaps = 4/604 (0%)

30 Query: 1 NCGIVGVGNRNATDILMQLEKLEYRGYSAGIFVANANQTNLKSQVRIADIRAKIGI 60  
 NCGIVG +G +A +IL+GLEKLEYRGYSAGI VAN ++ K GRADIAR +  
 Sbjct: 1 NCGIVGYIGQLDAKEILKLEKLEYRGYSAGI LANVEQSHVFKENGRADIAREVVER 60

Query: 61 DVAGSTGIGHTRWATHQSSTEDNAMPHTSOTGRFVLVHNGVNIENYLHIKTEFLAGHDFG 120  
 +V GIGETRWATHG+ + NAHPR S GRF LVHNGVNIENY +K E+L + K  
 Sbjct: 61 NVBEAKAGIGHTRWATHGEPSTLANHPHQSAIGRFTLVHNGVNIENYQLKQETLQVVELKS 120

35 Query: 121 QTCTEIAVHLIGKFVEEDKLSVLEAFKKSLSIIBGSYAPALMDSQATTIYVARNKSFLL 180  
 TYTE+ V +I +FV L EAF+K+L++++GSYA AL D+ +TI+VARNKSFLL  
 Sbjct: 121 DTDETVVQVIEQFVANG-LTEBAFRKTLTLKSGSYALAFNDNRETIYVARNKSFLL 179

40 Query: 181 IGLSGYNNVCSDAVAMIRETSEFMEIHDKELVILTKDKVTVITYDGKELIRDSYTAELD 240  
 +GLG+ +N+V SDMMAM++ T+E++E+ DKE+VI+T D+V + + DQ + R SY AELD  
 Sbjct: 180 VGLSGDTFNVASDAVAMQLVINEYVELMDKEMVITVDDQVVIKNDGDVITRASYTAELD 239

45 Query: 241 LSDIGKCTYPPFMYLKEIDQTPVNRQLISTYADETGRNVQVDPAITTSIQADRLYTLIAG 300  
 SDI NGTYP YMLKE DEQP VNR++I TY DE G + V I ++ EADR+YI+ G  
 Sbjct: 240 ASDIGKCTYPPHMYLKEIDQFVVMRKIIQTYQDENGKLVSPGIDAAVAASADRIYIIGQ 299

50 Query: 301 TSYHAGFATKNMLQLTDTVPELVASEGVMHPLLSKKMPFLSCSGSETADSRQVLNV 360  
 TSYHAG K +E + FVE+ VASE+ Y+MPLLSKKP+FT LSCSGSETADSR VLN+  
 Sbjct: 300 TSYHAGLNGKQYIEMMANVPEVHVASEFSTYNNMPLLSKKPLFTFLSCSGSETADSRVLNV 359

Query: 361 ANNMGIPSLTVNVPGSTLGRATYTMILIHAGPEIAVASTKAYTQAIALAPLAKAVCEA 420  
 A+G +LT+INVPGSTLGR EA YT+L+HAGPEIAVASTKAYTQAIA LA LA +  
 Sbjct: 360 VKALGRKALTINVPGSTLGRADYTLILHAGPEIAVASTKAYTQAIAVLAVLASVAADK 419

55 Query: 421 NGKQEAIDFNLVHLSVAQSIEATLSKKDLVAEKVQALATTNAPYITGRNDYTVAME 480  
 NG + F+LV EL + A ++EA +KD + + L +RNAP+TGRG DY+V +E  
 Sbjct: 420 NGIN--IGFDLVKSLGIANAMEALCDQKDEMEMLAREYLVTSNRNAPPIGRGLDYFVCE 477

60 Query: 481 AALKKEISYIQEGFANGELKHGTISLIEQTFVIALISSSOLVASHTGNIQEVVSRG 540  
 ALKKEISYIQ EGFA GELKHGTI+LIE+ TPV AL + + + S RGN++EVAARG  
 Sbjct: 478 GALKKEISYIQEGFANGELKHGTIALIEQTFVIALATQEHVNL-IRGNVKEVAARG 536

65 Query: 541 AHVLTVVESELDRGGDDIIVNKHVPFLAPIAWVITPQLIAYYASLQGLVDVKPRNLAKA 600

-1466-

A+ + +GLD D ++ +V+P LAP+ V+P QLIAYYA+L RG DVDKPRNLAK+  
 Sbjct: 537 ANTCIISLKGGLDADDFVLPEVNPALAFVSVVPLQLIAYYALRGCDVDKPRNLAKS 596

Query: 601 VTVE 604  
 VTVE  
 Sbjct: 597 VTVE 600

An alignment of the GAS and GBS proteins is shown below.

Identities = 500/604 (82%), Positives = 552/604 (90%)

Query: 1 MCGIVGVGNATDILIQLEKLEYRGYDSAGIFVVGDNKSLQVKGVRIRIAHQAVGD 60  
 MCGIVGVGN NATDIL+QGLEKLEYRGYDSAGIFV N++ L+KSVGRIA++AK+G  
 Sbjct: 1 MCGIVGVGNRNATDILANQLEKLEYRGYDSAGIFVANANQTNLIKSVGRITADLRKIGI 60

Query: 61 SVSGTGTIGTRWATHGKPTBGNAPHITSGSRFVLVHNGVNIENYLIQIKEYTLYTHNLRG 120  
 V+G+TGIGTRWATHG+ TE NAPHIT+GRFVLVHNGVNIENYL IK +L +H+ KG  
 Sbjct: 61 DVAGTGTIGTRWATHGQSTEDNAPHITSGTGRFVLVHNGVNIENYLHIKTEPLAGHDFKG 120

Query: 121 ETUTEIAIHLVHFVEEDNLSVLEAPFKALHIIEGSAFALIDSQADDTIYVAKNKSPLL 180  
 +TUTEIA+HL+ FVEED LSVLEAPFK+L IIEGSAFAL+DSQ DTIYVAKNKSPLL  
 Sbjct: 121 QTUTEIAVHLGKFVEEDKLSVLEAPKSLSIIEGSAFALMDSQATDTIYVAKNKSPLL 180

Query: 181 IGLGNGYNVCSDDAMAMIRETSEYMEIHDKELVIVKDKSVEVQYDGNVIERGSYTAELD 240  
 IGLG GYNVCSDDAMAMIRETSE+MEIHDKELVI+ KD V V DYDG + R SYTAELD  
 Sbjct: 181 IGLGEGYNVCSDDAMAMIRETSEYMEIHDKELVILITKDKVTVDYDGKELIRDSYTAELD 240

Query: 241 LSDIGKGTFFPYMLKEIDBQPTVMRLISTYANESGDMVNSDIKSVQEAERLYILAAG 300  
 LSDIGKGTFFPYMLKEIDBQPTVMR+LSTYA+E+G++ VD II S+QEAERLYILAAG  
 Sbjct: 241 LSDIGKGTFFPYMLKEIDBQPTVMRQLISTYADETGNVQVDPALITTSIQEAERLYILAAG 300

Query: 301 TSYHAGFAAKTMKRLTDTPVELGVSSGYNMPLLSKKKMPILLSSQSGETADSRQVLK 360  
 TSYHAGFA K M+E+LTDTPVELGV+SENGY+MPLLSKKKMPILLSSQSGETADSRQVLK  
 Sbjct: 301 TSYHAGFATQNMRLTDTTPVELGVASENGYHMPLLSKKMPILLSSQSGETADSRQVLK 360

Query: 361 ANWGIPISLTITNVPGSTLSREATYTMILHAGPEIAVASTKAYTAQVATLAFIAKAVGEA 420  
 AN WGIPISLT+TNVPGSTLSREATYTMILHAGPEIAVASTKAYTAQ+A LAFIAKAVGEA  
 Sbjct: 361 ANWGIPISLTITNVPGSTLSREATYTMILHAGPEIAVASTKAYTAQAIALAFIAKAVGEA 420

Query: 421 NGKAEKDFDLAHELSTVAQSEIATLSEKDVISEKVEQLLISTRNAPFYIQRGNDYVYME 480  
 NGK EA DF+LAHELST+VAQSEIATLSEKDV++EKV+ LL +TRNAPFYIQRGNDYV ME  
 Sbjct: 421 NGKQALDFDLAHELSTVAQSEIATLSEKDLVAEKVQALLATRNAPFYIQRGNDYVYME 480

Query: 481 AALKLKEISYIQCBFAAGELAKHOTISLIEEDTPVIALISADSTIAHTRGNIQRVSRG 540  
 AALKLKEISYIQ BFAAGELAKHOTISLIE+TPVIALIS+ +A+HTRGNIQRV +RG  
 Sbjct: 481 AALKLKEISYIQCBFAAGELAKHOTISLIEEDTPVIALISSQLVASHTRGNIQEVAARG 540

Query: 541 ANALLIIVEGELERBDDIIVNKVHPPLSAISMVPTQLIAYYASLQRLGVLDKPRNLAKA 600  
 A+ L +VBEGL+RBDDIIVNKVHPPL+ I+MVPTQLIAYYASLQRLGVLDKPRNLAKA  
 Sbjct: 541 AHNLVVEBGLERBDDIIVNKVHPPLAPIMVPTQLIAYYASLQRLGVLDKPRNLAKA 600

Query: 601 VTVE 604  
 VTVE  
 Sbjct: 601 VTVE 604

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1331

A DNA sequence (GBSx1415) was identified in *S. agalactiae* <SEQ ID 4081> which encodes the amino acid sequence <SEQ ID 4082>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have a cleavable N-term signal seq.



-1467-

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 9797> which encodes amino acid sequence <SEQ ID 9798> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAC44435 GB:U65000 type-I signal peptidase SpsB [Staphylococcus  
 aureus]  
 Identities = 62/185 (33%), Positives = 97/185 (51%), Gaps = 12/185 (6%)  
 Query: 10 VKRDFIRNIILALIALVLLILLRYFVFATFKVHKDATNSYFSGDVVVVN---RNRTFK 65  
 +K++ ++ II +A+IL++ F+ ++ ++ ++ +G+ V VN + +  
 15 Sbjct: 1 MKCELLEMIISIAVAFVILFIVGKFIVTPTTIKESMDPTLKDGERVAVNIIGYKTKGLE 60  
 Query: 66 YKDFIVYKVGKIF-YISRVIGEPNQKVRVMDILYLVDFKDEPYIERMKMAYSEKKDQG 124  
 + +V+ K Y+ RVIG P KV +D LY+N +DEPY+ N + K G  
 20 Sbjct: 61 KGVVVFHANRNDYVYKRVIGVPGDKVEYKNTLYLVGKQKDEPYL---NTNLGKQGD 116  
 Query: 125 MPFTSDFSVETL--TRNKESRVPEKGSYLVLNDNRQNKNDKRKGLIKEKIDRGVITFKVY 182  
 T F V+ L K + +PKG YLVL DNR+ DSR FGLI E I G ++F+ +  
 Sbjct: 117 Y-ITGTQVKDLFANPKSNVIPKGYKGLVLGDNREVSKDSRPFGLIDEDQIVGKVSFRFW 175  
 25 Query: 183 PLSEF 187  
 P SEF  
 Sbjct: 176 PFSEF 180

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4083> which encodes the amino acid sequence <SEQ ID 4084>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -14.22 Transmembrane 10 = 26 ( 4 - 34)

35 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6689 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 99/185 (53%), Positives = 130/185 (69%)  
 Query: 9 MVKRDIFIRNIILALIALVLLILLRYFVFATFKVHKDATNSYFSGDVVVVNRNRTPKYKD 68  
 MVKRDIFIRNI+L LI ++ ILLR FVP+TPKV + N+Y +GD+V +N PKYKD  
 45 Sbjct: 1 MVKRDIFIRNIILLIIVIGAILLRIFVFSTFKVSPETANTYLKSGDLVTIKNIQPKYKD 60  
 Query: 69 FIVYKVGKIFYISRVIGEPNQKVRVMDILYLVDFKDEPYIERMKMAYSEKKDQMPFT 128  
 F+Y+VGK Y+SRVI V MDI YLN++ + Y+ERMK Y +T  
 50 Sbjct: 61 FVVYKVGKIDYSRVIAVBGDSVTYMDIPIYLNMVBSQYLIERMKMAYLHAPFGTLYT 120  
 Query: 129 SDPSVETLIRNKESRVPEKGSYLVLNDNRQNKNDKRKGLIKEKIDRGVITFKVYPLSEFG 188  
 DF+V T+T +K +VPG YL+LNDNR+N NDSR+PGLI I+G++TP+V PLS+FG  
 Sbjct: 121 DDFVATITADKYCKVPEKGYLLNDNRNTNDSRRFGILNASQIGKLVTFRVLPISDFG 180  
 55 Query: 189 FTASE 193  
 F E  
 Sbjct: 181 FVEVE 185

60 A related GBS gene <SEQ ID 8789> and protein <SEQ ID 8790> were also identified. Analysis of this protein sequence reveals the following:

-1468-

```

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: 10.13
GVH: Signal Score (-7.5): 0.45
 Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 3.82 threshold: 0.0
PERIPHERAL Likelihood = 3.82 69
modified ALOM score = -1.26

```

\*\*\* Reasoning Step: 3

----- Final Results -----

```

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

36.0/59.9% over 165aa

Bacillus caldolyticus  
EGAD|24914| signal peptidase i Insert characterized

ORF00169 (364 - 867 of 1179)

EGAD|24914|25718(15 - 180 of 182) signal peptidase i {Bacillus caldolyticus}

%Match = 11.9

```
%Identity = 35.9 %Similarity = 59.9
```

Matches = 60 Mismatches = 61 Conservative Sub.s = 40

```

312 342 372 402 432 462 483 510
L*GIDIMEKRLGVVMKRFIRMIILALIVALLILLLYLVFAATVKVHKDATSYPNSGDVVVNR--NRPRTKQ-DPI
 : : : : || : : || : : : : : : : : : : : : : : : : : :
 VTKQKEKGRGRVPMFVAVCVATLELVESNVVVEGKSMPFLESNIIIVNKLSDYDIPRRFDI
 10 20 30 40 50 60

537 567 597 627 657 687 717 747
VTKVKGKIP-YISKVIGPMQKQVMDVIMLIVKDFDSFIRNKNAYSEKKDGMFFTSYVETLTRNKESVKGYSY
 :
VPHANGKGYDYSYVIGLFGRLAYMDVKKVQKVDSPVLPKQ--KLLQRL--KQDFLEVT--GKTRVPGD
 80 90 100 110 120 130 140

```

[illegible]

SEQ ID 8790 (GBS7) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 4; MW 46kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 4; MW 21kDa). The GBS7-His fusion product was purified (Figure 189, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 262), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1332

A DNA sequence (GBSx1416) was identified in *S. galactiae* <SEQ ID 4085> which encodes the amino acid sequence <SEQ ID 4086>. Analysis of this protein sequence reveals the following:

Possible site: 54

```
>>> Seems to have no N-terminal signal sequence
```

-1469-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1099 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9795> which encodes amino acid sequence <SEQ ID 9796> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25804 GB:AF172173 pyruvate kinase [Streptococcus thermophilus]  
 Identities = 413/500 (82%), Positives = 451/500 (89%)

Query: 1 MNKRKIVATLGPVAFRGGKKFGESEGYWESLDVEASAKIAQLIKEGANVFRNFSHG 60  
 MNKRKIVATLGPVAFRGGKKFGESEGYWESLDVEASAKIAQLIKEGANVFRNFSHG 60  
 Sbjct: 1 MNKRKIVATLGPVAFRGGKKFGESEGYWESLDVEASAKIAQLIKEGANVFRNFSHG 60

Query: 61 DHABQGRMATVRKASEIAQQKVGFLDTKGPEIRTELFEDGADFHSSYTTGKLRVATQ 120  
 +HABQGR MAT VR AS IAQQKVGFLDTKGPEIRTELF E A ++Y TG ++R+ATQ  
 Sbjct: 61 NHABQGRMDVVRMAESIAQQKVGFLDTKGPEIRTELFEDGADFEYATYKGEQIRIATQ 120

Query: 121 GIKSTPEVIALNVAAGLDIFDDVEVGRQLVDDGKLGLTVPKDKDTREFFVVENEGLI 180  
 G+KST +VIALNVA G LDIFDDVEVGRQ+LVDDGKLG L V KD + REF V VENEG+I  
 Sbjct: 121 GLKSTRDVIALNVAAGLDIFDDVEVGRQLVDDGKLGRLVVDKDAKREFIVEVENEGII 180

Query: 181 GKQGVNIPYTKIPFPALAEEDNADIRFGLQGGINFIAISFVRTAKDVENRAICEETGN 240  
 KQGVNIPYTKIPFPALAEEDNADIRFGLQG+NFI AISFVRTAKDV EVRAICEETGN  
 Sbjct: 181 AKQGVNIPYTKIPFPALAEEDNADIRFGLQGGINFIAISFVRTAKDVQEVRAICEETGN 240

Query: 241 GHVLLAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVFPFVYQRMIIITKVNAAGK 300  
 GHVLL AKIENQQGIDNIDEIIEAADGIMIARGDMGIEVFPFVYQRMIIITKVNAAGK  
 Sbjct: 241 GHVLLAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVFPFVYQRMIIITKVNAAGK 300

Query: 301 AVITATNMLETMDKPRATRSVSDVFNVIDGTDATMLSGESANGKYPVESVKTATID 360  
 V+ITATNMLETMDKPRATRSVSDVFNVIDGTDATMLSGESANG YPVESVKTATID  
 Sbjct: 301 IIVTATNMLETMDKPRATRSVSDVFNVIDGTDATMLSGESANGYPVESVKTATID 360

Query: 361 KNAQTLNLEYGRLDSSAFFRNKDKDVIAAVKDATSHMDILVVTITETGNTARAISKFR 420  
 KNAQTLN LEYGR+SS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R  
 Sbjct: 361 KNAQTLNLEYGRLDSSAFFRNKDKDVIAAVKDATSHMDILVVTITETGNTARAISKFR 420

Query: 421 PDADILAVTFDEKQVRLINWGVIPVLADKFASTDMDFEVABRVALEAGFVSGENIVI 480  
 P+ADI A+TFDE Q+SLM+NWGVIPV+ + P+STDMDFEVABRVALE+G VESGENIVI  
 Sbjct: 421 PDAIDILAVTFDELTKQSLMIAWGVIPVVTETFSSTDMDFEVABRVALESSLVESGENIVI 480

Query: 481 VAGVFPVGTGGTINMRKTVK 500  
 VAGVFPV+G T+TMR+RTVK  
 Sbjct: 481 VAGVFPVSGNTINMRIRTVK 500

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4087> which encodes the amino acid sequence <SEQ ID 4088>. Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0915 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

RGD motif: 272-274

The protein has homology with the following sequences in the databases:

>GP:AAF25804 GB:AF172173 pyruvate kinase [Streptococcus thermophilus]  
 Identities = 404/500 (80%), Positives = 457/500 (90%)

-1470-

Query: 1 MNKRVKIVATLGPVAVETRGKKYGEDGYWAQQLDVEBSAKKIAELIAGANVFRFNFSG 60  
 Sbjet: 1 MNKRVKIVATLGPVAVETRGKKYGEDGYWSEKLDPTASAKNIAQLIEBGANVFRFNFSG 60

Query: 61 DIKEQGDQRMATVRLABETARQKVGFLDITKGPEMTELFPADDAKEPSYVTGKEKRVATTQ 120  
 +H EQG+RM VR+AE IA QKVGFLDITKGPE+RTLEF DAK+Y+Y TGE+IR+AT Q  
 Sbjet: 61 NEAQGGRRMDVVRMAESLAGQKVGFLDITKGPEIRTELFPDGAKEYAYKTGEQIRLRIATQ 120

Query: 121 GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDDGKLGKLVKDQIATRQFIVEVEGDII 180  
 G++STRDVIALNVAG+LDI+D+VEVG +L+HDGKLG+V+D+K R+FIVEVEGDII  
 Sbjet: 121 GLKSTRDVIALNVAGALDI+DDVEVGKQLVDDGKLGKLVVDKDAEKREIFIVEVEGDII 180

Query: 181 AKQKGVNIPNTKIPFPALASRDNADIRPGLQGGLNFIAISFVRTAKDVVEVREICKSTGN 240  
 AKQKGVNIP TKIPFPALASRDNADIRPGLQGQ+NFIAISFVRTAKDV+EVV IC ETGN  
 Sbjet: 181 AKQKGVNIPYTKIPFPALASRDNADIRPGLQGGLNFIAISFVRTAKDVQEVRAICESTGN 240

Query: 241 DHVQLFAKIENQQGIDNLDIIEAADGIMIARGDMGIEVFPFEMVVPVQKMIITKVNAAGK 300  
 HV+L AKIENQQGIDN+DEIIIEAADGIMIARGDMGIEVFPFEMVVP+QMIITKVNAAGK  
 Sbjet: 241 GHVQLFAKIENQQGIDNLDIIEAADGIMIARGDMGIEVFPFEMVVPVQKMIITKVNAAGK 300

Query: 301 AVITATNMLETMTKPRATRSVSDVFNVAIDGIDATMLSGESANGKYPVESVKTMTID 360  
 V+TATNMLETMTKPRATRSVSDVFNVAIDGIDATMLSGESANG YPVESVKTMTID  
 Sbjet: 301 IVVATNMLETMTKPRATRSVSDVFNVAIDGIDATMLSGESANGKYPVESVKTMTID 360

Query: 361 RNAQTLLNEYGRLDSSAFPRNNKTDVIAASVKDATHSMIDKLWVTTITETGNRAISKFR 420  
 +NAQTLL ETKGL+SS F R++ T+V+ASVKDAT+SM I+L+V +TE+GNTA I +R  
 Sbjet: 361 RNAQTLLNEYGRLDSSSTFDRSSNTEVVAASVKDATHSMIQLIVALTESGNTASLIDTYR 420

Query: 421 PDADILAVIFDEKQVRLALMINWGVIPVLAEKPASTDGMFEVAEKVAEAGLQVSGENIVI 480  
 P+ADI A+TFDR Q++LM+NWGVIPV+ E P+STDGMFEVAEKVA+E+GLV+SGENIVI  
 Sbjet: 421 PEADIMAITFDLTKQLSLMLNWGVIPVVTETPSSTDGMFEVAEKVALESGLVESGENIVI 480

Query: 481 VAGVPGVGCGTNRKTRVK 500  
 VAGVPG+G TNR+RTRVK  
 Sbjet: 481 VAGVPGSGNTNRKTRVK 500

An alignment of the GAS and GBS proteins is shown below.

Identities = 440/500 (88%), Positives = 462/500 (92%)

Query: 1 MNKRVKIVATLGPVAVETRGKKYGEDGYWAQQLDVEBSAKKIAELIAGANVFRFNFSG 60  
 MNKRVKIVATLGPVAVETRGKKYGEDGYW LOVE SA+KIA+LI+ GANVFRFNFSG 60  
 Sbjet: 1 MNKRVKIVATLGPVAVETRGKKYGEDGYWAQQLDVEBSAKKIAELIAGANVFRFNFSG 60

Query: 61 DHAEQGMATVRLABETARQKVGFLDITKGPEIRTELFPDGAKEPSYVTGKEKRVATTQ 120  
 DH EQG RMAVTR AEEIA QKVGFLDITKGPE+RTLEF D A SY TG K+RVAT Q  
 Sbjet: 61 DHAEQGMATVRLABETARQKVGFLDITKGPEMTELFPADDAKEPSYVTGKEKRVATTQ 120

Query: 121 GIKSTPEVIALNVAGSLDIYDEVEVGHTILIDDGKLGKLVKDQIATRQFIVEVEGDII 180  
 GI+ST +VIALNVAG LDI+D+VEVG IL+HDGKLG V KD TR+P V VENG+I  
 Sbjet: 121 GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDDGKLGKLVKDQIATRQFIVEVEGDII 180

Query: 181 GKQKGVNIPYTKIPFPALASRDNADIRPGLQGGLNFIAISFVRTAKDVVEVREICKSTGN 240  
 GKQKGVNIP TKIPFPALASRDNADIRPGLQGGLNFIAISFVRTAKDV EVR IC ETGN  
 Sbjet: 181 AKQKGVNIPNTKIPFPALASRDNADIRPGLQGGLNFIAISFVRTAKDVVEVREICKSTGN 240

Query: 241 GHVQLFAKIENQQGIDNLDIIEAADGIMIARGDMGIEVFPFEMVVPVQKMIITKVNAAGK 300  
 HV+LFAKIENQQGIDN+DEIIIEAADGIMIARGDMGIEVFPFEMVVP+QMIITKVNAAGK  
 Sbjet: 241 DHVQLFAKIENQQGIDNLDIIEAADGIMIARGDMGIEVFPFEMVVPVQKMIITKVNAAGK 300

Query: 301 AVITATNMLETMTKPRATRSVSDVFNVAIDGIDATMLSGESANGKYPVESVKTMTID 360  
 AVITATNMLETMTKPRATRSVSDVFNVAIDGIDATMLSGESANGKYPVESVKTMTID  
 Sbjet: 301 AVITATNMLETMTKPRATRSVSDVFNVAIDGIDATMLSGESANGKYPVESVKTMTID 360

Query: 361 RNAQTLLNEYGRLDSSAFPRNNKTDVIAASVKDATHSMIDKLWVTTITETGNRAISKFR 420  
 +NAQTLLNEYGRLDSSAFPR NTDVIAASVKDATHSMIDKLWVTTITETGNRAISKFR  
 Sbjet: 361 RNAQTLLNEYGRLDSSAFPRNNKTDVIAASVKDATHSMIDKLWVTTITETGNRAISKFR 420

-1471-

- Query: 421 PDADILAVTFDEKVRSLMINWGVIPVLADKPASTDDMFEVASKVALEAGFVSGDNIVI 480  
 PDADILAVTFDEKVRSLMINWGVIPVLA+KPASTDDMFEVASKVA+EAG V+SGDNIVI  
 Sbjct: 421 PDADILAVTFDEKVRSLMINWGVIPVLARKPASTDDMFEVASKVALEAGFVSGDNIVI 480
- Query: 481 VAGVFPVGTGGTNTMRVRTVK 500  
 VAGVFPVGTGGTNTMRVRTVK  
 Sbjct: 481 VAGVFPVGTGGTNTMRVRTVK 500
- 10 A related GBS gene <SEQ ID 8791> and protein <SEQ ID 8792> were also identified. Analysis of this protein sequence reveals the following:
- Belongs to Glycolysis/gluconeogenesis pathway. Proteins belonging to this metabolic pathway have been experimentally detected on the surface of *Streptococci*.
- The protein has homology with the following sequences in the databases:
- 15 >GP|6708106|gb|AAP25804.1|AF172173\_2|AF172173 pyruvate kinase  
 {*Streptococcus thermophilus*}
- Score = 821 bits (2098), Expect = 0.0  
 Identities = 412/500 (82%), Positives = 450/500 (89%)
- Query: 1 MNKRVKIVATLGPVAVEPRGGKKGSGYWGSELDVZASAKIAQLIKEGANVFRFPNPSHG 60  
 MNKRVKIVATLGPVAVEPRGGKKGSGYWGSELDVZASAKIAQLIKEGANVFRFPNPSHG  
 Sbjct: 1 MNKRVKIVATLGPVAVEIRGGKKGSGDYWSEKLDVZASAKIAQLIKEGANVFRFPNPSHG 60
- Query: 61 DHAEQGARMATVRKAEIAGQKVGFLDTKGPEIRTELFDGADPFHSYTTGTKLKRVATQ 120  
 +HAEQG RM VR AE IAGQKVGFLDTKGPEIRTELE A ++Y TG ++R+ATKQ  
 Sbjct: 61 NHAEQGERMDVVRKAEIAGQKVGFLDTKGPEIRTELEFDGADPFHSYTTGTKLKRVATQ 120
- Query: 121 GIKSTPEVIALNVAGSLIDFDDVEVGKQLVDDGKLGTLVFAKDKCTREFEVVVENDGLI 180  
 G+KST +VIALNVAG LDI+DDVEVGKQ+LVDDGKLG L V KD + REF V VENDG+I  
 Sbjct: 121 GIKSTRDVIALNVAGSLIDFDDVEVGKQLVDDGKLGTLVVDKDKAREFIVEVENDGII 180
- Query: 181 GKQKGVNIPTTKIPFPALAEARNADIRFGLQGGLNFIAISFVRKADQVNEVRAICBETGX 240  
 KQKGVNIPTTKIPFPALAEARNADIRFGLQGGLNFIAISFVRKADQVNEVRAICBETGX  
 Sbjct: 181 AKQKGVNIPTTKIPFPALAEARNADIRFGLQGGLNFIAISFVRKADQVNEVRAICBETGN 240
- Query: 241 GHVKLAKIENQQGIIINIDRIIEAADGINIARGMGIEVPFEMVPVYQRMIIKVNAGK 300  
 GHVKL AKIENQQGIIINIDRIIEAADGINIARGMGIEVPFEMVPVYQRMIIKVNAGK  
 Sbjct: 241 GHVKLLAKIENQQGIIINIDRIIEAADGINIARGMGIEVPFEMVPVYQRMIIKVNAGK 300
- Query: 301 AVITATNMLETMTDKPRATSEVSDVFNVAIDGTDATMLSGESANGKYPVSRVTMATID 360  
 V+TATNMLETMT+KPRATSEVSDVFNVAIDGTDATMLSGESANG YPVSRVTMATII  
 Sbjct: 301 IVVATNMLETMTDKPRATSEVSDVFNVAIDGTDATMLSGESANGKYPVSRVTMATIH 360
- Query: 361 KNAQTLLNEYGRLLSSA+FRNNKTDVIAASAVKDAHSMDIKLVVITITGTNRAISKFR 420  
 KNAQTLL EYGRLLSS F R++ T+V+ASAVKDA+SM I+L+V +TE+GNTA I +R  
 Sbjct: 361 KNAQTLLNEYGRLLNSST+FRNSNTVEVAVKDAHSMDIKLVVITITGTNRAISKFR 420
- Query: 421 PDADILAVTFDEKVRSLMINWGVIPVLADKPASTDDMFEVASKVALEAGFVSGDNIVI 480  
 P+ADI A+TFDE Q+SLM+NWGVIPV+ P+STDMMFEVASKVALE+G VESGDNIVI  
 Sbjct: 421 PDADILAVTFDEKVRSLMINWGVIPVVTETPSSTDMMFEVASKVALESGDNIVI 480
- Query: 481 VAGVFPVGTGGTNTMRVRTVK 500  
 VAGVFPVGTGGTNTMRVRTVK  
 Sbjct: 481 VAGVFPVGTGGTNTMRVRTVK 500

SEQ ID 8792 (GBS330) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 5; MW 59kDa).

GBS330-His was purified as shown in Figure 213, lane 6.

- 60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1472-

**Example 1333**

A DNA sequence (GBSx1417) was identified in *S.galactiae* <SEQ ID 4089> which encodes the amino acid sequence <SEQ ID 4090>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0632 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF25803 GB:AF172173 phosphofructokinase [Streptococcus thermophilus]
Identities = 270/337 (80%), Positives = 302/337 (89%), Gaps = 1/337 (0%)

Query: 1 MKRIAVLTSGGDAPGMAAIRAVVRKAISEGMEVYGINQGYGMVTDGIPFLDANSVGD 60
MKRIAVLTSGGDAPGMAAIRAVVRKAISEGMEVYGINQGYGMVTDGIPFLDANSVGD 60
Sbjct: 1 MKRIAVLTSGGDAPGMAAIRAVVRKAISEGMEVYGINQGYGMVTDGIPFLDANSVGD 60

Query: 61 INRGGTFLESARYPEFALEGGQLKGIEQLKHGIBGVVVGIDGDSYHGAMRLTEHGFPAV 120
++RGSTFLSARYPEFALEGGQLKGIEQLKHGIBGVVVGIDGDSYHGAMRLTEHGFPAV 120
Sbjct: 61 LSRGTFVLSARYPEFALEGGQLKGIEQLKHGIBGVVVGIDGDSYHGAMRLTEHGFPAV 120

Query: 121 GLPGTIINDIVGTDYTIIGFDATAVATAENLDRLDTSASHNRTFVVEVMGRNAGDIALWS 180
GLPGTIINDIVGTDYTIIGFDATAVATA E LD++DT+ SH RTFVVEVMGRNAGDIALWS 180
Sbjct: 121 GLPGTIINDIVGTDYTIIGFDATAVATAEALDKIQDTAFSHRTFVVEVMGRNAGDIALWA 180

Query: 181 GIAAGADQIIVPEEFNIDEVVSINVRAGVAG-IKHQIIVLAEVGMSGDEFAKIMGAAGD 239
GIA+GADQIIVPEE++I+EVV V+GY+G K H IIVLAEVGM +EFA MK AGD
Sbjct: 181 GIAAGADQIIVPEEYDINEVVRKVEGYESSEKSHHIIIVLAEVGMAEFAKIMKAGD 240

Query: 240 DSOLRVNLGHLLRGSGPTARDRVLASMGAYAVOLLKRGGLAVGHNEEMVSPILG 299
SOLR TNLGH++RGSGPTARDRVLAS MGA+AV LLKGG GG+AVG+HNE++VESPIILG
Sbjct: 241 TSQLRATNLGHVIRGSGPTARDRVLASMGHAAVDLLKGGIGGAVGHNEQLVESPIILG 300

Query: 300 LAEEGALFSLTDEGKIIVNPHKADRLAALNRDLAN 336
AEEGALFSLT++GKI+VNNPHKA L A LNR LAN
Sbjct: 301 AEEGALFSLTDEGKIIVNPHKADRLAALNRDLAN 337

```

Proteins in the glycolysis/gluconeogenesis pathway have been experimentally detected on the surface of *Streptococci*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4091> which encodes the amino acid sequence <SEQ ID 4092>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0632 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 274/336 (81%), Positives = 306/336 (90%), Gaps = 1/336 (0%)

Query: 1 MKRIAVLTSGGDAPGMAAIRAVVRKAISEGMEVYGINQGYGMVTDGIPFLDANSVGD 60
MKRIAVLTSGGDAPGMAAIRAVVRKAISEGMEVYGINQGYGMVTDGIPFLDANSVGD 60
Sbjct: 1 MKRIAVLTSGGDAPGMAAIRAVVRKAISEGMEVYGINQGYGMVTDGIPFLDANSVGD 60

Query: 61 INRGGTFLESARYPEFALEGGQLKGIEQLKHGIBGVVVGIDGDSYHGAMRLTEHGFPAV 120

```

-1473-

```

I+RGSTFL SARYPEFA+LEEQI GIEQLKKHGIGVVVIGGDSYHGAMRLTSHGFTAV
Sbjct: 61 ISRGSTFLYSARYPEFAQLSEQLAGIEQLKKHGIGVVVIGGDSYHGAMRLTSHGFTAV 120

5 Query: 121 GLPGTIDNDIVGDTYTTIGFDTAVATAVENLDRDTSASHNRITFVVEVMRAGDIALKS 180
G+PGTIDNDI GTDTTIGFDTAV TAVE +D+LRDTS+SH RTFVVEVMRAGDIALK+
Sbjct: 121 GIPGTIDNDIAGDTYTTIGFDTAVNTAVEALDKLRDTSSSHGRITFVVEVMRAGDIALKA 180

Query: 181 GIAAGADQIIVPEKEFNIDRVSVNRAGYA-AGKHQIIVLAGVMSGDKFAKTMKANGD 239
GIA+GADQIIVPESEF+I+V S ++ + GK+H IIVLAGVMSG+ FA+ +K AGD
10 Sbjct: 181 GIASGADQIIVPEEFDEKVASTIQYDFEHGKRNNHIIIVLAGVMSGGEFAQKLEKAGD 240

Query: 240 DDLRVNLNGLHLLRGSSPTARDVLAARMGAYAVQLLKEGSGGLAVGVHNEEMVESPIIG 299
SDLRVNLNGLH+LRGGSPTARDRV+AS MG++AV+LLK+G+GGLAVG+HNRE+VESPIIG
Sbjct: 241 KDLRVNLNGLHLLRGSSPTARDVLAWSMGSHAVELLDKGGGLAVGVHNEELVESPIIG 300

15 Query: 300 LASEGALFSLTDEGKIIVNNPHKADLRILAALNRDLA 335
ASEGALFSLT+EGKI+VNNPHKA L AALNR L+
Sbjct: 301 TASEGALFSLTEGKIIVNNPHKARLDFAALNRSLA 336

```

- 20 SEQ ID 4090 (GBS313) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 5; MW 41kDa).

GBS313-His was purified as shown in Figure 204, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 1334

A DNA sequence (GBSx1418) was identified in *S. agalactiae* <SEQ ID 4093> which encodes the amino acid sequence <SEQ ID 4094>. This protein is predicted to be DNA polymerase III alpha subunit (dnaE). Analysis of this protein sequence reveals the following:

```

Possible site: 55
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1446 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

There is also homology to SEQ ID 4096.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 40 Example 1335

A DNA sequence (GBSx1419) was identified in *S. agalactiae* <SEQ ID 4097> which encodes the amino acid sequence <SEQ ID 4098>. This protein is predicted to be YHCF (farR). Analysis of this protein sequence reveals the following:

```

Possible site: 52
45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3316 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-1474-

>GP:BA04102 GB:AP001508 transcriptional regulator (GntR family)  
[Bacillus halodurans]  
Identities = 51/116 (43%), Positives = 79/116 (67%)

5 Query: 5 FNEKSPIYSQIAEHIMQIVSQEIKSGDQLPTVRELAQAGVNPNTMQRAFTLEKRGV 64  
P+ PIY Q+AB +K QIV E++ G++LP+VR++ BA VHEVT+QR + ELE +V  
Sbjct: 5 FHSSEPIYQLAERVRQIVRGELRLGKLPVSRMGIEANVNPNTVQRTYRELEGLKIV 64

10 Query: 65 FSORTSGRPFVTEENLLIGKIRQQVAKAELATFVNMKKIGYKLEIDTEVALDHFKE 120  
S+R G FVIED ++ IR+Q+ + E++ FV M+++GY +EI L+ ++ E  
Sbjct: 65 ESKRGQHFVTEDEBQLALREPMKETEISHFVQGMREMGYSINRIQAGLESYLTE 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4099> which encodes the amino acid sequence <SEQ ID 4100>. Analysis of this protein sequence reveals the following:

15 Possible site: 25  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
20 bacterial cytoplasm --- Certainty=0.2075 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/120 (66%), Positives = 100/120 (82%)

25 Query: 1 MAWEFNEKSPIYSQIAEHIMQIVSQEIKSGDQLPTVRELAQAGVNPNTMQRAFTLEK 60  
M+W+F EKSPFIY+QIA+H+ MQI+SQEIKSGDQLPTVRE A+ AGVNPNTMQRAFTLEK  
Sbjct: 1 MSWKEEKSPIYAQIAHVMQIVSQEIKSGDQLPTVREYAEIAGVNPNTMQRAFTLEK 60

30 Query: 61 BGMVFSORTSGRPFVTEENLLIGKIRQQVAKAELATFVNMKKIGYKLEIDTEVALDHFKE 120  
BGMVFSORT+GRPFVT+D LI + R++++ +EL +F+ NM K+G+ EI L F+KE  
Sbjct: 61 BGMVFSORTAGRPVTDQKLIARKKRELAISELESFITNMTKMGFSHTEIIPVLTSFKE 120

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1336

A DNA sequence (GBSx1420) was identified in *S.agalactiae* <SEQ ID 4101> which encodes the amino acid sequence <SEQ ID 4102>. This protein is predicted to be ABC transporter, ATP-binding protein (yhcG). Analysis of this protein sequence reveals the following:

40 Possible site: 26  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
45 bacterial cytoplasm --- Certainty=0.2757 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12735 GB:Z99108 similar to glycine betaine/L-proline  
transport [Bacillus subtilis]  
Identities = 87/228 (38%), Positives = 150/228 (65%), Gaps = 1/228 (0%)

50 Query: 5 LQLHHVTKYKHHTAVNDVTSIETPGKILGLGPNKSGKFTIIMKINGLLQPOKGDIVD 64  
++L EV+KKY +HDAVNDV+++ +G+I GL+GPNKSGK+T +KM+ GLL P G +D  
55 Sbjct: 3 IKLEHVSKKYGRHTAVNDVSIETLSSGRITGLGPNKSGKSTLLKMAAGLFPETGIVKD 62

Query: 65 GYRPSVETKIIISYLPDTSYLQENMKIKDVTLFEDFYNDPDSKVAYQLPEDLNINPRR 124  
+ + E + +YL + + +KD+V ++ + DF ++ Y+L ++ INP ++



-1475-

Sbjct: 63 EEQVIREMVRQTAYLITKELDMFYPHFTVKIMVNFYQSQFPDHFTEQVYKLLNEMQLNPEKK 122

Query: 125 LKMLSGNKKKQVLILVMSRKARLYILDEPIGGVDPAAARDYILKTIISNYSNDAS-VLIS 183  
+K LSGN+ +++++L ++R+A + +LDEP G+DP RD I+ +++S + V+I+

Sbjct: 123 LKMLSGNKRGLKIVLALARRADVILILDEPFGSLDPMVRDSIVNSLVSYIDFSGQIVVIA 182

Query: 184 THILSDIEPIILDEVIFLKRGEIDLQGNADDLRREHNCSDIALFRERFK 231

TH I +IE +LDEVI L GE Q +D+RE+ S+ F+ + +

Sbjct: 183 THIEIDETLLDEVIILANSEKVAQREVDIRKQGMVSLQWFKSKME 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4103> which encodes the amino acid sequence <SEQ ID 4104>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1983 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/231 (74%), Positives = 200/231 (86%)

Query: 1 MTQLQLLHVHTKKYKHTAVNDVTVSIPGKIIIGLLPGNSGKTTIIRKINGLLQPDGKD 60  
M LQLLHV+K Y + A++D+T++IP GRIIGLLPGNSGKTTI+K+INGLLQ+K+G+

Sbjct: 1 MAHLQLLHVHVSREKKAIDDLTITIPIGRIIGLLPGNSGKTTILKILKINGLLQPDGKD 60

Query: 61 IVIDGYRPSVETKIIISYLPDTSYLQNMKIKDVTYLFEDFYNDPDSKVAQLFEDILNL 120

IVIDGYRP VETKIIISYLPDT+YL ENM+IND++ F DFY+DEF A L DL L+

Sbjct: 61 IVIDGYRPSVETKIIISYLPDTSYLQNMKIKDVTYLFEDFYNDPDSKVAQLFEDILNL 120

Query: 121 FRERKLNLSGNKKKQVLILVMSRKARLYILDEPIGGVDPAAARDYILKTIISNYSNDASV 180

P +R K LSGNKKKQVLILVMSRKARLY+LDEPIGGVDPAAARDYILKTII++Y +ASV

Sbjct: 121 PEDRFKTLNSGNKKKQVLILVMSRKARLYILDEPIGGVDPAAARDYILKTIISNYSNDASV 180

Query: 181 LISTHLSDIEPIILDEVIFLKRGEIDLQGNADDLRREHNCSDIALFRERFK 231

+ISTHLSDIEPIILDEVIFLK+G + L GNADDLR+E+ SID+LFFE +K

Sbjct: 181 IISTHLSDIEPIILDEVIFLKRGLPLSGNADDLRQETQCSIDSLFRFTYK 231

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1337

A DNA sequence (GBSx1421) was identified in *S.agalactiae* <SEQ ID 4105> which encodes the amino acid sequence <SEQ ID 4106>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -15.39 Transmembrane 120 - 136 ( 103 - 146)

INTEGRAL Likelihood = -9.98 Transmembrane 55 - 71 ( 47 - 79)

INTEGRAL Likelihood = -9.45 Transmembrane 22 - 38 ( 15 - 43)

INTEGRAL Likelihood = -6.05 Transmembrane 192 - 208 ( 187 - 218)

INTEGRAL Likelihood = -4.94 Transmembrane 230 - 246 ( 228 - 253)

INTEGRAL Likelihood = -4.78 Transmembrane 157 - 173 ( 155 - 175)

INTEGRAL Likelihood = -1.44 Transmembrane 103 - 119 ( 103 - 119)

----- Final Results -----

bacterial membrane --- Certainty=0.7156 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEP database.

-1476-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4107> which encodes the amino acid sequence <SEQ ID 4108>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
5 INTEGRAL Likelihood = -11.52 Transmembrane 190 - 206 { 187 - 215 }
 INTEGRAL Likelihood = -10.67 Transmembrane 121 - 137 { 104 - 141 }
 INTEGRAL Likelihood = -5.73 Transmembrane 63 - 79 { 59 - 82 }
 INTEGRAL Likelihood = -4.83 Transmembrane 158 - 174 { 156 - 181 }
10 INTEGRAL Likelihood = -1.38 Transmembrane 232 - 248 { 232 - 248 }
 INTEGRAL Likelihood = -0.85 Transmembrane 104 - 120 { 104 - 120 }

----- Final Results -----
 bacterial membrane --- Certainty=0.5607 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 116/267 (43%), Positives = 165/267 (61%), Gaps = 13/267 (4%)

20 Query: 1 MFGKLLKYEKSVGKGYLTINAAVLIVSIIILGLVLKALG-----GNFSTDTNSTSAQIFT 55
 MFGKLLKYEK +S+GKGY LNA V+ ++ IL +K G F TN ++
 Sbjct: 1 MFGKLLKYEKFRSIGWYFALNAFVIAIAALLSFTIKLFAQNSDGLFGVLTN---KNLP 56

25 Query: 56 IILVLLAMVIGSLLSTLAIILIRFYSNIFGRQQYTLTLPLVTTNQCISKLASLWIS 115
 + L L +I+GSLSTL IIIIRF ++FG +GILTTLTLPV ++QII SKLAS + S
 Sbjct: 57 IILGLTFGSLIAGSLSTLIIILIRFYSNIFGRQQYTLTLPLVTTNQCISKLASLWIS 116

30 Query: 116 IFNIFIVIIIGILLVILPLVGIQGFVVAFFEYKIISSNAFLFIAPFLSVAGTLLIYL 175
 +FN I+ I +VI+P+ I + + F +K+ N +AY LS LLIYL
 Sbjct: 117 VFNTIIAFAIAIVIVMFNINELLEGFNSFNDYFINDMLTVLAVLSTFTSILLIYL 176

35 Query: 176 SIAVQQLFNRKRVLMGIVSYFGISLLITFLTLIIDSIFHIDLENSHANA-TFSQPVLLY- 233
 SI++QQLF+N+R LM ++YF + +LI+ + S EI N+ A++ F++ +Y
 Sbjct: 177 SISIQQFNSNRGLMAFIAYFVILIVISVAATTVHS--HFNINTSADSPFPTQKTIYL 234

40 Query: 234 NIIAVSIVEIAIFYMLTHSIKKYKLNQ 260
 IL +E+ +FY+ T+ IIK KLN+Q
 Sbjct: 235 LILEQFIEMIMFYLATNFTIKKLNQ 261

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1338

A DNA sequence (GBSx1422) was identified in *S.agalactiae* <SEQ ID 4109> which encodes the amino acid sequence <SEQ ID 4110>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5890 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein is similar to ORF24 from *S.faecalis*.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1477-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1339

A DNA sequence (GBSx1423) was identified in *S.agalactiae* <SEQ ID 4111> which encodes the amino acid sequence <SEQ ID 4112>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to ORF23 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1340

A DNA sequence (GBSx1424) was identified in *S.agalactiae* <SEQ ID 4113> which encodes the amino acid sequence <SEQ ID 4114>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4256(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to ORF22 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1341

A DNA sequence (GBSx1425) was identified in *S.agalactiae* <SEQ ID 4115> which encodes the amino acid sequence <SEQ ID 4116>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-13.37 Transmembrane 62 - 78 (55 - 84)
 INTEGRAL Likelihood = -8.44 Transmembrane 19 - 35 (14 - 41)

----- Final Results -----
bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to ORF21 from *S.faecalis*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4117> which encodes the amino acid sequence <SEQ ID 4118>. Analysis of this protein sequence reveals the following:

Possible site: 37  
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2444(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/236 (22%), Positives = 95/236 (39%), Gaps = 12/236 (5%)

Query: 204 KDGKRLRLMKNVWEYDKLPHMLIAGGTGGGKTYFILTLIEALLHTDSKLYILDPKN---- 259  
+ GK+ ++K+ DK H IAG +G GK Y LT ++L S L I+DPK  
Sbjct: 14 CQGIKIFVIKHFEMLNLDKGSWAIAGNSGSGKPY-ALTYFLSVLKPKKSLIIIDPKFDTPS 72

Query: 260 --ADLADLGSMANVYYRKEDLLSCIETFYEMMKRSEEMKQMKNYKTGKNYAYLGLPAH 317  
A + + + + K D + S + + + + + + L +  
Subject: 73 CWARENKIAVIHPVENHKSDFVSQVNEOLNOCATLIQKRAILYDNPNHOFTHLTI--- 129

Query: 318 FLIFDEYVAFMEMLGTKENIAVMNKLKQIVMLGRQAGFFLILACQRPDAKYLGDGIRDQF 377  
+ DE +A E + A + L QI +LG L L QR D + +R+Q  
Sbjct: 130 --VIDEVLALSEGYNKNIEAFSSLISQALLGHATKIHLFLGSRQFDHNTIPISVREOL 187

Query: 378 NFRVALGRMSENGYGMMPGSDVQKDFLLKRIKGRGYVDVGTSVISEFYTLPVPKGY 433  
N + +G +++ +F + + G G + V + S PL+ Y  
Sbjct: 188 NVLLQIGNINOKTTFLLFPDLDPGEIVPTGHTGTGIIQVVDNEHSYQVLPLLCPTY 243

SEQ ID 4116 (GBS109d) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 8 & 9; MW 71kDa) and in Figure 184 (lane 2; MW 71kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 11; MW 46kDa), Figure 128 (lane 4; MW 46kDa) and Figure 179 (lane 7; MW 46kDa). GBS109d-His was purified as shown in Figure 232 (lanes 7 & 8). GBS109d-GST was purified as shown in Figure 236, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1342

A DNA sequence (GBSx1426) was identified in *S.galactiae* <SEQ ID 4119> which encodes the amino acid sequence <SEQ ID 4120>. Analysis of this protein sequence reveals the following:

Possible site: 22  
>>> Seems to have a cleavable N-term signal seq.

```
----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1343**

A DNA sequence (GBSx1427) was identified in *S.agalactiae* <SEQ ID 4121> which encodes the amino acid sequence <SEQ ID 4122>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4469 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9793> which encodes amino acid sequence <SEQ ID 9794> was also identified.

The protein is similar to ORF20 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1344**

A DNA sequence (GBSx1428) was identified in *S.agalactiae* <SEQ ID 4123> which encodes the amino acid sequence <SEQ ID 4124>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1367 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1345**

A DNA sequence (GBSx1429) was identified in *S.agalactiae* <SEQ ID 4125> which encodes the amino acid sequence <SEQ ID 4126>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.77 Transmembrane 39 - 55 (34 - 64)
 INTEGRAL Likelihood = -6.32 Transmembrane 16 - 32 (10 - 35)

----- Final Results -----
 bacterial membrane --- Certainty=0.5310 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein is similar to ORF19 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

-1480-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1346

A DNA sequence (GBSx1430) was identified in *S.agalactiae* <SEQ ID 4127> which encodes the amino acid sequence <SEQ ID 4128>. This protein is predicted to be antirestriction protein. Analysis of this protein sequence reveals the following:

Possible site: 22  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2918(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to ORF18 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1347

A DNA sequence (GBSx1431) was identified in *S.agalactiae* <SEQ ID 4129> which encodes the amino acid sequence <SEQ ID 4130>. Analysis of this protein sequence reveals the following:

Possible site: 27  
>>> Seems to have a cleavable N-term signal seq.  
INTEGRAL Likelihood = -3.61 Transmembrane 75 - 91 ( 72 - 94)

----- Final Results -----  
bacterial membrane --- Certainty=0.2444(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to ORF17 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8793> and protein <SEQ ID 8794> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 4  
MoG: Discrim Score: -7.12  
GvH: Signal Score (-7.5): -2.52  
Possible site: 43  
>>> Seems to have no N-terminal signal sequence  
ALOM program count: 1 value: -3.61 threshold: 0.0  
INTEGRAL Likelihood = -3.61 Transmembrane 37 - 53 ( 34 - 56)  
PERIPHERAL Likelihood = 3.66 58  
modified ALOM score: 1.22

\*\*\* Reasoning Step: 3

----- Final Results -----  
bacterial membrane --- Certainty=0.2444(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1481-

100.0/100.0% cver 167aa

Enterococcus faecalis

EGAD[14977] hypothetical protein Insert characterized  
 GP[532550][gb|AAB60016.1||U09422 ORF17 Insert characterized

5

ORF00720(187 - 690 of 990)

EGAD[14977][15011(1 - 168 of 168) hypothetical protein {Enterococcus faecalis}  
 GP[532550][gb|AAB60016.1||U09422 ORF17 {Enterococcus faecalis}

10

%Match = 50.3

%Identity = 100.0 %Similarity = 100.0

Matches = 168 Mismatches = 0 Conservative Sub.s = 0

15

```

120 150 180 210 240 270 300 330
L*AKYQLVFKTILIIKPMVGI*TPQERLSQPIMSGFLKSSIKSVGTTLLADFLFYGVQAQATPIFYERIDYMKKIRSYTSI
 |||
 M3PLKSSIKSVGTTLLADFLFYGVQAQATPIFYERIDYMKKIRSYTSI
 10 20 30 40

```

20

```

360 390 420 450 480 510 540 570
WSVEKVLYSINDFRLFPPIITFTMTWVVSFLFAVMILGNLPFLSMIEGAFLKYFGIPVAFWFMSTKTPDGGKKPYGFLKS
|||
WSVEKVLYSINDFRLFPPIITFTMTWVVSFLFAVMILGNLPFLSMIEGAFLKYFGIPVAFWFMSTKTPDGGKKPYGFLKS
60 70 80 90 100 110 120

```

25

```

600 630 660 690 720 750 780 810
VIAYALRPKLTAYAGKKVTLGRNQFORAITAVRSEFYGISN*IH*KQSRLE*RMGMCLL*ACSLQLLISKSRNTESA*P
|||
VIAYALRPKLTAYAGKKVTLGRNQFORAITAVRSEFYGISN
140 150 160

```

30

SEQ ID 8794 (GBS223) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 7; MW 18kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 35 Example 1348

A DNA sequence (GBSx1432) was identified in *S.agalactiae* <SEQ ID 4131> which encodes the amino acid sequence <SEQ ID 4132>. Analysis of this protein sequence reveals the following:

Possible site: 37

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

40

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4292 (Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

45

A related GBS nucleic acid sequence <SEQ ID 9791> which encodes amino acid sequence <SEQ ID 9792> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50

### Example 1349

A DNA sequence (GBSx1433) was identified in *S.agalactiae* <SEQ ID 4133> which encodes the amino acid sequence <SEQ ID 4134>. Analysis of this protein sequence reveals the following:

-1482-

Possible site: 16

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -6.21 | Transmembrane | 350 - 366 ( 345 - 368) |
| INTEGRAL | Likelihood = -0.32 | Transmembrane | 171 - 187 ( 171 - 188) |

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.3484 (Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1350**

- 15 A DNA sequence (GBSx1434) was identified in *S.agalactiae* <SEQ ID 4135> which encodes the amino acid sequence <SEQ ID 4136>. Analysis of this protein sequence reveals the following:

Possible site: 45

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.30 | Transmembrane | 154 - 170 ( 148 - 177) |
| INTEGRAL | Likelihood = -10.30 | Transmembrane | 21 - 37 ( 17 - 50)     |
| INTEGRAL | Likelihood = -10.03 | Transmembrane | 320 - 336 ( 316 - 367) |
| INTEGRAL | Likelihood = -7.43  | Transmembrane | 346 - 362 ( 337 - 367) |
| INTEGRAL | Likelihood = -7.01  | Transmembrane | 186 - 202 ( 180 - 206) |
| INTEGRAL | Likelihood = -5.35  | Transmembrane | 411 - 427 ( 404 - 430) |
| INTEGRAL | Likelihood = -1.17  | Transmembrane | 386 - 402 ( 386 - 402) |

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5118 (Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1351**

- 35 A DNA sequence (GBSx1436) was identified in *S.agalactiae* <SEQ ID 4137> which encodes the amino acid sequence <SEQ ID 4138>. Analysis of this protein sequence reveals the following:

Possible site: 14

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial cytoplasm | --- | Certainty=0.6306 (Affirmative) | < succ> |
| bacterial membrane  | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-1483-

**Example 1352**

A DNA sequence (GBSx1437) was identified in *S.agalactiae* <SEQ ID 4139> which encodes the amino acid sequence <SEQ ID 4140>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2973 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1353**

A DNA sequence (GBSx1438) was identified in *S.agalactiae* <SEQ ID 4141> which encodes the amino acid sequence <SEQ ID 4142>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3382 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

There is also homology to SEQ ID 4144.

A related GBS gene <SEQ ID 8795> and protein <SEQ ID 8796> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 3
McG: Discrim Score: 11.12
GVH: Signal Score (-7.5): 0.27
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 4.19 threshold: 0.0
PERIPHERAL Likelihood = 4.19 69
modified ALOM score: -1.34

*** Reasoning Step: 3

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

100.0/100.0% over 332aa

*Enterococcus faecalis*

```

EGAD|36209| hypothetical protein Insert characterized
GP|532547|gb|AAB60019.1||U09422 ORF14 Insert characterized

ORF00727(301 - 1299 of 1599)
EGAD|36209|37602(1 - 333 of 333) hypothetical protein {Enterococcus
faecalis|GP|532547|gb|AAB60019.1||U09422 ORF14 {Enterococcus faecalis}

```

-1484-

%Match = 61.7  
 %Identity = 100.0 %Similarity = 100.0  
 Matches = 333 Mismatches = 0 Conservative Sub.s = 0

```

5 249 279 309 339 369 399 429 459
 CSEKSTTTTKYK*TTNQNRHH*EER*ETMKLKLTVIGGSGGLFLMWFSILLFVAILFSDSQSGTSMITHYGGVNVSAEVLAH
 MKLKLTVIGGSGGLFLMWFSILLFVAILFSDSQSGTSMITHYGGVNVSAEVLAH
 10 20 30 40 50

10 489 519 549 579 609 639 669 699
 KPMVEKYAKYGVVEYVNILLATIQVESGGTAEADVMSSESLGLLPPMSLSTESIKQGVKYPSELLASSERLSVDLESVI
 KPMVEKYAKYGVVEYVNILLATIQVESGGTAEADVMSSESLGLLPPMSLSTESIKQGVKYPSELLASSERLSVDLESVI
 70 80 90 100 110 120 130

15 729 759 789 819 849 879 909 939
 QSYNYGGGFLGYVANRGNKYTFELAQSFSEKYSGGEEKVSYNPIAIPINGGWRNYGNMPTVQLVYIQLVITTFDDDTVQ
 QSYNYGGGFLGYVANRGNKYTFELAQSFSEKYSGGEEKVSYNPIAIPINGGWRNYGNMPTVQLVYIQLVITTFDDDTVQ
 150 160 170 180 190 200 210

20 969 999 1029 1059 1089 1119 1149 1179
 AIMDEALKYEGWRVYVGGASPTTDFDCSGLTQWTYTGKAGINLPRTAQQQDVTQHIPLSEAQAGDLFFHSTINAGSYIT
 AIMDEALKYEGWRVYVGGASPTTDFDCSGLTQWTYTGKAGINLPRTAQQQDVTQHIPLSEAQAGDLFFHSTINAGSYIT
 230 240 250 260 270 280 290

30 1209 1239 1269 1299 1329 1359 1389 1419
 HVGIIYLGNNRMFHAGDPIGYADLTSPFWQQLVVGAGRIKQ*ERKI***NLEKIRIKKNRYQRKNLVSIRSLIKRL*LP
 HVGIIYLGNNRMFHAGDPIGYADLTSPFWQQLVVGAGRIKQ
 310 320 330

```

35 SEQ ID 8796 (GBS155) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 10; MW 38kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 7; MW 62kDa).

The GBS155-GST fusion product was purified (Figure 111; see also Figure 198, lane 74) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoreactive on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1354

45 A DNA sequence (GBSx1439) was identified in *S. agalactiae* <SEQ ID 4145> which encodes the amino acid sequence <SEQ ID 4146>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.60 Transmembrane 37 - 53 (35 - 55)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.4439 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

55

-1485-

A related GBS nucleic acid sequence <SEQ ID 9789> which encodes amino acid sequence <SEQ ID 9790> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1355

A DNA sequence (GBSx1440) was identified in *S.agalactiae* <SEQ ID 4147> which encodes the amino acid sequence <SEQ ID 4148>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -0.00 Transmembrane 391 - 407 ( 391 - 407)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9787> which encodes amino acid sequence <SEQ ID 9788> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4149> which encodes the amino acid sequence <SEQ ID 4150>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2027(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 183/669 (27%), Positives = 305/669 (45%), Gaps = 63/669 (9%)

Query: 7 KIINIGVLAHVADAGKTTLTSLYNSGAITELQSVDKGTTRTDNILLERQGITITQGIT 66  
 K NIG++AVHDACKTT TE +LY +G I ++G +G ++ D E++RGITI + T  
 Sbjct: 9 KTRNIGIMAHVDAGKTTTTERILYYTGKIKIGETHEGASQMDWMEQQRGITITSAT 68

Query: 67 SFQWENTKVNIIIDTPGHMDFLAEVYRSLSLVLDGAILLISAQDVQCACTRILPHALKRGI 126  
 + QN+ +VNIIIDTPGH+DF EV RSL VLDGA+ ++ ++ GV+ QT ++ + G+  
 Sbjct: 69 TAGNDGRHVNIIIDTPGHVDFITVQRSLRVLDGAVTVLDGSGSVPEPTETVWRQATEYGV 128

Query: 127 PTIFPFINKIDQNGIDLSIVYQDIKRLSARI-----VIKQKVELYN 168  
 P I F NK+D+ G D Q + ++L A +IK K E+Y N  
 Sbjct: 129 PRIVFANKMDKIGADFLYSVCLHDLRQANAHPIQLFIGAEDDFGIIIDLIKMKARIYN 188

Query: 169 MCVTNFTS---BQW-----DTVIEGNDLLEKYMGRSLALELEQBSIRF 213  
 T+ E E++ + V E ++DL+ Ky+ G+ + EL  
 Sbjct: 189 DLGTDILREDDIPESYLEQAQCYRRLILBAVATDIEDLMMKYLBSGRTINDELIAGIRKAT 248

Query: 214 HNCSLFPVYHSGAKNNIGIDNLKVI-----TNKPYSSSTURGPGR---L 254  
 N FVY GSA N G+ +++ + N + P+  
 Sbjct: 249 INVEFFVLQCSAFLKNGVQLMDLAVLAYLPSPLDI PAI KGVNPDTDAREERPASDEFP 308

Query: 255 CGNVKIEYTKKQRLAYIRLYSGVILHRDSSVRVSEKEI---KVTMYTINSIGELCKI 310  
 FKI RL + R+YSGVI+ V + K K ++ +M+ + E I  
 Sbjct: 309 AALAFKIMTDFVGRILTPFRVYSGVINSQVYVMYTSKGRKRIRGLQMDANSRQ---I 365

-1486-

Query: 311 DRAYSGEIVILQN-EPLKIANSVLGDTKLLPQRKKIENPHLLQTTVEPSKPKQREMLLDA 369  
 + Y+G+I + L D K + IE P P+Q VEP ++ + A  
 5 Sbjct: 366 KTVYAGDIAAAGVGLKDTTIGDSLTDEKAKVILEIEVPEPVIQLMVEPKSKADQDKMGVA 425

Query: 370 LLEISDSDFLLRYVVDSTHEIILSFLGKVMQEVISALLQKQYHVEIKLKEPTVIYME-- 427  
 L ++++ DP R + T E +++ +G++ ++V+ ++ ++ VE + P V Y E  
 Sbjct: 426 LQKLAREDFTPRVETNVETGETVIAGMGLHLGLVLDVRMKKRPKVRANVGAPQVSYRETF 485

10 Query: 428 RPLKNAEYTHIEVFPNPFNASIGLSVSPFLGSGMQYBSSVSLGYINQSPQNAVMBEIR 487  
 R A + + + + +P G G ++E+++ G + + F AV +G+  
 Sbjct: 486 RASTQARGFPFKQSGGKSQFGDWIEFTPNBEGKGFEFMAIVGGVVPREPIPAVERGLI 545

15 Query: 488 YGCBQG-LYGMNVTDCKICFKYGLYSPVSTPADFRMLAPIVLQVLKKGATELLEFPYLS 546  
 G L G + + D K G Y+ S+ F++ A + L++ K A +LEP +  
 Sbjct: 546 ESMANGVLAGYPMVDVKACLYDGSYHVDGSETAFKLAASLALKEAASQAQPAILEMML 605

20 Query: 547 FKITYAPQEYLSRAYNDAPKYCANIVDTQLQNEVILSGEIPARCIQEYRSDLTFFPTNGRS 606  
 I AP++ L + + N I+ +P + Y + L T GR  
 Sbjct: 606 VTTIAPEDNLGDMGHVTRRRGVDGMEAHGNSQIVRAIVPLAEMFGYATVLRASATQGRG 665

Query: 607 VCLTELKGY 615  
 + Y  
 Sbjct: 666 TPKMVFDRHY 674

25

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1356

A DNA sequence (GBSx1441) was identified in *S.agalactiae* <SEQ ID 4151> which encodes the amino acid sequence <SEQ ID 4152>. Analysis of this protein sequence reveals the following:

Possible site: 33  
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2530 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1357

A DNA sequence (GBSx1442) was identified in *S.agalactiae* <SEQ ID 4153> which encodes the amino acid sequence <SEQ ID 4154>. Analysis of this protein sequence reveals the following:

45 Possible site: 18  
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1487-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1358

A DNA sequence (GBSx1443) was identified in *S.agalactiae* <SEQ ID 4155> which encodes the amino acid sequence <SEQ ID 4156>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1630 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1359

A DNA sequence (GBSx1444) was identified in *S.agalactiae* <SEQ ID 4157> which encodes the amino acid sequence <SEQ ID 4158>. This protein is predicted to be excisionase-related protein. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4481 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein is similar to transposon Tn916 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1360

A DNA sequence (GBSx1445) was identified in *S.agalactiae* <SEQ ID 4159> which encodes the amino acid sequence <SEQ ID 4160>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4626 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein is similar the Tn1545 integrase from *S.pneumoniae* and to SEQ ID 578.

-1488-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1361

A DNA sequence (GBSx1446) was identified in *Sagalactiae* <SEQ ID 4161> which encodes the amino acid sequence <SEQ ID 4162>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.72 | Transmembrane | 18 - 34 ( 13 - 41)     |
| INTEGRAL | Likelihood = -6.10  | Transmembrane | 58 - 74 ( 55 - 79)     |
| INTEGRAL | Likelihood = -5.04  | Transmembrane | 97 - 113 ( 90 - 116)   |
| INTEGRAL | Likelihood = -1.81  | Transmembrane | 78 - 94 ( 78 - 94)     |
| INTEGRAL | Likelihood = -0.85  | Transmembrane | 145 - 161 ( 145 - 161) |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5288(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database.

>GP:AACT74820 GB:AB000270 orf, hypothetical protein [Escherichia coli K12]  
Identities = 43/174 (24%), Positives = 84/174 (47%), Gaps = 9/174 (5%)

Query: 24 LIATLVLVVLYKL-----GIIDNSNELKDLVHKYEPWGPIMFIVAQIVQVFPVPG3 77

Sbjct: 24 LFACLIFALVITAIHAGFLFDLLTULPHLQTLIRGSGFFGYSLYILLFIATLL-LLPGS 82

Query: 78 VTTVAGFLIFGPTLGFPIYNYGIIIGSVILWLVKFKYGRKFVLLFM-DQKTFDKYKSKLE 136

Sbjct: 63 ILVIAAGGIVGFLGTLGLSLIAATLASSCSFLARWLGRDLLKIVGHSNTPQALEKGIA 142

Query: 137 TSGYEKFFIFCMASPISPADIMMVTGLSNMSIKRFVTIIMITKPSIIGSYL 190

Sbjct: 143 RRGID-FLIITLILPLFPINIQNYAGLTTFAPWYTLISALTLPQIVITVFM 195

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4163> which encodes the amino acid sequence <SEQ ID 4164>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                      |
|----------|--------------------|---------------|----------------------|
| INTEGRAL | Likelihood = -4.30 | Transmembrane | 8 - 24 ( 6 - 29)     |
| INTEGRAL | Likelihood = -0.80 | Transmembrane | 57 - 73 ( 57 - 73)   |
| INTEGRAL | Likelihood = -0.00 | Transmembrane | 86 - 102 ( 86 - 102) |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.2720(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 85/114 (74%), Positives = 101/114 (88%)

Query: 89 PTLGFYINYGIIGSVILWLVKFKYGRKFVLLFMDQKTFDKYKSKLETSGYEKFFIFCM 148

Sbjct: 3 PVTGTYINYGIIGSIALFLVKTIVYGRKFILLFVNDKTFYKYERLRLTGPYKELIFCM 62

Query: 149 ASPISPADINVMITGLSNMSIKRFVTIIMITKPSIIGSYLWYIGDGLKNFL 202

Sbjct: 63 ASPVSPADINVMITGLSNMSIKRFVTILLITKPSIIGSYLITPGKDVISNFI 116

-1489-

There is also homology to SEQ ID 1728.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 1362

A DNA sequence (GBSx1447) was identified in *S.agalactiae* <SEQ ID 4165> which encodes the amino acid sequence <SEQ ID 4166>. This protein is predicted to be chlorAMphenicol acetyltransferase (cat). Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial cytoplasm | --- | Certainty=0.4725 (Affirmative) | < succ> |
| bacterial membrane  | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA86871 GB:U19459 VAT B [Staphylococcus aureus]  
 Identities = 57/130 (43%), Positives = 81/130 (61%), Gaps = 4/130 (3%)

20 Query: 57 IGAPCSIAQNV--ITGLNHPTDHTTNPFIIYKSRGFNEDRADLIDKIKGKVIIGND 114  
 IG FC+IA+ + + G NH + ITT PF G+ + L D G ++GND  
 Sbjct: 65 IGKFCALAEIGIEFIMGANHRMNSITTYFF-NIMGNM-EKATPSLEDLPGKGDVTGVND 122

25 Query: 115 VMIGNTVITLPSVTIGNGAIIGAGSVITKIDIPDYAVAGTPAKIIKYPSEETILNAS 174  
 VMIG NVT++P + IG+GAI+ A SV+TKD+P Y ++ G P++IIK RF +E I L  
 Sbjct: 123 VMIGQNVTVMPGIQIGDGAIVANSVVTKDVPFPIIIGNPSRIIKRPFEDLIDYLLQI 182

30 Query: 175 QWNNWSDEAI 184  
 +WW+NS + I  
 Sbjct: 183 KQWWDNSAQKI 192

There is also homology to SEQ ID 1944.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1363

A DNA sequence (GBSx1448) was identified in *S.agalactiae* <SEQ ID 4167> which encodes the amino acid sequence <SEQ ID 4168>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial cytoplasm | --- | Certainty=0.2398 (Affirmative) | < succ> |
| bacterial membrane  | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 1364

A DNA sequence (GBSx1449) was identified in *S. agalactiae* <SEQ ID 4169> which encodes the amino acid sequence <SEQ ID 4170>. This protein is predicted to be cation-transporting P-ATPase PacL. Analysis of this protein sequence reveals the following:

5      Possible site: 34  
      >>> Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -9.18 | Transmembrane | 873 - 889 ( 866 - 894) |
| INTEGRAL | Likelihood = -8.39 | Transmembrane | 257 - 273 ( 251 - 276) |
| INTEGRAL | Likelihood = -5.95 | Transmembrane | 67 - 83 ( 65 - 88)     |
| INTEGRAL | Likelihood = -5.41 | Transmembrane | 282 - 298 ( 281 - 301) |
| INTEGRAL | Likelihood = -1.65 | Transmembrane | 90 - 106 ( 89 - 107)   |
| INTEGRAL | Likelihood = -0.48 | Transmembrane | 737 - 753 ( 736 - 753) |
| INTEGRAL | Likelihood = -0.00 | Transmembrane | 898 - 914 ( 898 - 914) |

10

15      ----- Final Results -----  
          bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>  
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20      A related GBS nucleic acid sequence <SEQ ID 10963> which encodes amino acid sequence <SEQ ID 10964> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AB85991 GB:AB000912 cation-transporting P-ATPase PacL  
          [Methanothermobacter thermoautotrophicus]

25      Identities = 409/922 (44%), Positives = 609/922 (65%), Gaps = 22/922 (2%)

Query: 10    TNTRFAKEELEEVFEELGTTQGGILSDSEVAVRQKKYGLNLLSEVKBOSIILLFLKNFTSL 69  
          T T    + E+EEV + L T+ GL +E R K + G N L EVK+ +ILFL N ++

30      Sbjct: 4    TMTAIYELEVEEVQLRLETSSEGLDQKAEKKIKLGHGPNKLEEVKRRPILLLFLNLYNV 63

Query: 70    MAILLVGGFVAIVSNSLEGLAIWMVNVINGI+PSPIQEYRASQATQALEHMLPSYSRVL 129  
          +A+LLR+    ++ ++ + +L +AI MV +IN +FSF QEY A +A +AL+ +LP +V+

35      Sbjct: 64    LALLMLIAAILSFITGNYLAVAIWMVILINALPSFWQYEASKAABALKNILFVMVKVI 123

Query: 130    RKSGEEKILSEQLVPGDIVLIEEQRISADGRLIKTTDLQVNSALTYGSMPIYKDSNVE 189  
          R E I +    +V GDI+++EEGD + AD R+++ +L+V+ SALTYGS P+ K S+

40      Sbjct: 124    RASKEVLIPADVYVHGDIIILEEGSDTFEADARILIEHNLRFVDASALTYGSFVRCVSHPV 183

Query: 190    NDQSKTILIECDNVFAGTTVSSGQATMVVTAIGMCTQFGQIADLTQMKISEKSPLEQLD 249  
          + +    + I+ +N++FAGT V+G3+ V A G T+F +IA LTQ ++ E SPLQR++

45      Sbjct: 184    RE-ALNYIDTENILFAGTQVTSQGRAAVFATGRDTEFSRIATLTQVEVRESEPLQRQIS 242

Query: 250    RLITQIISITVTGIIIFPLAATFFVKEPVSKSFYALGMIVAPIEGHLLPTVTLSLMAV 309  
          + I +++ +G+I FL + V+ P+ +FFA+G++VA +PEGLLP+VTLSLA +

50      Sbjct: 243    LAARIIGALAVAMGVILFLVNLVYIKLPLETAFIAGIAMVAVPEGLLPSVTLSLAASA 302

Query: 310    QRMKEHALVKLLSSVETLGATSVICSDIKCTH+QNEMTVNHLWQNKSKSYQVTLGYAPE 369  
          ++N+EA+ALNK+LSSVETLG+T++IC+DKTCTH+ EMTV +W K +VTG GY PE

55      Sbjct: 303    RKQARENALVKRLSSVETLSSTIIITCTDKTCTLRGEMTVRKINIPHKVILEVTSGGYRPE 362

Query: 370    QQLFEGDNICPQNSDRGDELKILIRFAHLCSNAQVLPPNDRSTTYTFLGDPTFACINVL 429  
          QQ LP G+ +    + D +L+ L+R A C++ ++    ++VLGD TE L V

60      Sbjct: 363    QQPLFRGEPV--SHRDMASKILIKRAATPCNDSALI---HEEGSEWSVLDTSEGALLVAA 417

Query: 430    EKSQINQIQRNKKFAPRIKELPFDSVRKRMTHIHLQGDCKKKISITWGAPKHEILDSDY 489  
          EK G + +    K PR+ ELFPDS RK MT+H G K+++ KGAPK+A L6+

65      Sbjct: 418    EKLQPDAAEELKAMPRITELPFDSRRKSMSTIHEKSG---KRVAVKGAPKXIIGLSER 473

Query: 490    VLSDQKVIPLNKEEENKIQLANDTFADKGLRLVLSVYQSDIEGSPKSEWTQENLKEGHVPI 549  
          + DG+V L+ +S+ +I +D A GLRVLA +Y ++ E +E+ +V+

70      Sbjct: 474    ISVDGRVRLHADEKRIIGIHDEMASKGLRLVLAFLAYRELPE-DLVRVDQGEVERDVLV 532



-1491-

Query: 550 GLIAMSDDPRBGVRAIDKCHASIRIIMTGDYGLTASIAKNIIGIRNDQAKVISGLE 609  
 G+ AM DPREGV+RA++ C A IRIIM+TDYGLTA +JA+ IGI+ + ++I G R  
 Sbjct: 533 GMAJMDHDPREGVKEAVESCKTAGIRIIMTGDYGLTAEATAREIGIVBG+ECRIIKOKE 591

5 Query: 610 LSEVTDSDQLKKELSGE--VVFARVAFHQKIRVVITLQEMGEVAVVATGDGVNDAPALKKSD 667  
 L++ D++L+ L+ E ++FAR FE K R+ ++L++ E+VA+TGDGVNDAPALA+K+D  
 Sbjct: 592 LDKLQVTLSEGLHGLARERNLIFARAVFPHKRIASVLESDSHIVAMTGDGVNDAPALRKAD 651

10 Query: 668 IGVAWGVTGTDVAKESADMLITDHFASIVHVAHEGRAVYQNIKKPLTYIFMSNTPHAPV 727  
 IGVAWG+GTDVAKE+AD+L DD+FAVIV AV EGR VY+NI+KF+TYIF+ T E VP  
 Sbjct: 652 IGVAWG-SGTDVAKESADIVLADNFAVIVAVREGRTVYENIRKRTITYIFSHETAIEIV 710

15 Query: 728 SAFFLFSKGFPLPLTVAQILANDLCTIMLPALGLGVEPPETVDVNNRPPRRITRILLKG 787  
 F + IFLP+T+MQILA+DLGTD LPAL LG PH+DVM PPR ++RL++  
 Sbjct: 711 --FIMMWLFSPLPITIMQILADLGTDTLPALALGRSLFSSDVNKLPPRAPSRLIARE 768

20 Query: 788 LLIKSFMLVGTIESVLAMGQFFWAIYLYGNF--TFVANGIPREATTMTILGALISQ 844  
 ++++ ++L+ GTIR+ L M +F Y G + A+ Y ATT+ I+ +Q  
 Sbjct: 769 VILRGYLTGTITREALIMATYELVLY--SGMWLPQQLSADSLYMRATLVVFAIGVAVG 826

25 Query: 845 IGWPNFSRTSYQSIALSIFGNKLINFGILIMEILAFVLVAVFLPHNLFWTASLGSHVYL 904  
 +G ++S+T S + N+ I G++ I L+++Y+P +F TA G+ W  
 Sbjct: 827 LGWLLSSQTLRSSALBGLLRNRWILAGMVFALSVMLLVITYLEPLQPIGTAPPGILEWF 886

30 Query: 905 YLISCFPIVIGLDEYKLFSSR 926  
 LI I+ DK+RK R  
 Sbjct: 887 LILLPFPIVFLTDKMRKPIQR 908

There is also homology to SEQ ID 4172.

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1365

A DNA sequence (GBSx1450) was identified in *S. agalactiae* <SEQ ID 4173> which encodes the amino acid sequence <SEQ ID 4174>. Analysis of this protein sequence reveals the following:

35 Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 40 bacterial cytoplasm --- Certainty=0.3740 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB46979 GB: AJ243482 CSRA protein [Enterococcus faecalis]  
 45 Identities = 85/132 (64%), Positives = 105/132 (79%)

Query: 2 KETQEELQRIGHAYQVQTQNSATHEAFTGKYDDFPFREGIYVDIVSGEVLFSLLDKPQSG 61  
 K T+REL+Q + Y VTO +ATE F+G+YDDF+++GIYVDIVSGE LFSLLDK+ +G  
 Sbjct: 3 KPTREELKQTLDTLQYAVTQENATPERFSGKYDDFYQDGIYVDIVSGEPLFSSLDKYDAG 62

50 Query: 62 CGWPAFSKPIENRMVNTNQHSHGMHRISVRSRQADSHLGHVFNDPVDAAGGLRYCINSA 121  
 CGWP+P+KPIE R V D SHGMHR+SVRS++ADSHLGHV F DGP+ GGLRYCIN+A  
 Sbjct: 63 CGWPSPTKPIEKGKGVKADFSHGMHRVSVRSQADSHLGHVFTDGLQGGGLRYCINPA 122

55 Query: 122 ALDPFIPYDQAK 133  
 AL F+P + K  
 Sbjct: 123 ALRFVVFVADLEK 134

- 60 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4175> which encodes the amino acid sequence <SEQ ID 4176>. Analysis of this protein sequence reveals the following:

-1492-

Possible site: 24  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.3692 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10       Identities = 109/142 (76%), Positives = 126/142 (87%)

Query: 3   ETQBELRQRIGHAYQVQTSNATHAFTGKYDDFFEEGIVYDIVSGEVLPSLDDKFKQSGC 62  
           ET +RL+QRIG +Y+VTQ++ATE FTG+YD+FFE+GIYVDIVSGEVLPSLDDKF SGC  
 15   Sbjct: 2   ETSDLELQRIIGDLSSVITQHAATESPFTGEYDNFFKGIYVDIVSGEVLPSLDDKFNSSGC 61

Query: 63   GNPAPSKPIENRMVITNHQDHSKGMIRIEVRSRQADSHLGHVFNEDGVPDAGGLRYCINSAA 122  
           GNPAPSKPIENRMVITNH D S+GM R+EV+SR+A SHLGHVF+DGP +AGGLRYCINSAA  
 15   Sbjct: 62   GNPAPSKPIENRMVITNHDDSSYGMRRVEVKSRKAGSHLGHVFSLGPKEAGGLRYCINSAA 121

20   Query: 123 LDFIFYPDQMAKRGYGYLSLFD 144  
           L FIFYPQM K GY +L+LFD  
           Sbjct: 122 LKFIYPDQMEKEGTQAQMLTLD 143

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 25   vaccines or diagnostics.

### Example 1366

A DNA sequence (GBSx1451) was identified in *S. agalactiae* <SEQ ID 4177> which encodes the amino acid sequence <SEQ ID 4178>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30                   bacterial cytoplasm --- Certainty=0.1674 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 35                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05127 GB:AP001511 unknown (Bacillus halodurans)

40       Identities = 48/152 (31%), Positives = 77/152 (50%), Gaps = 1/152 (0%)

Query: 1   MIRRAKEKDLFDIAELLEQIIMLHHEVRPDIFHTRGSKFSKEQLKMLIDESKPIFVYES 60  
           +IR A +D ++A L Q+ H + R DIF + + + + E + V+  
 40   Sbjct: 2   IIREATVQDYEEVARKLHTQVHEARVKERGDIFRSNEPTLNPSFFQAAYQGEKSTVLIVVD 61

Query: 61   DEGKVVAHLFIQLQEKRDLPK -KSFKTYIYDDLCDIEDEVRGQGIQQLMDFARQYAKKHG 119  
           + K+ A+ + L + LP + KT+YI DLC+DE RG IG+ + + Y K H  
 45   Sbjct: 62   EREKIGAYSVIHLVQTPLLPTMQRCITVYISDLCVDETRGGGIGRLIFRALISYGAHQ 121

Query: 120   CYNITLVNVDNQRAPSVFKLGPKQQTQMB 151  
           I L+V+ + N RA +FY LG + Q+ MR  
 50   Sbjct: 122 VDAIELDVYDNDRAKAFYHSLGMRQCKQIMR 153

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 55   vaccines or diagnostics.

**Example 1367**

A DNA sequence (GBSx1452) was identified in *S. agalactiae* <SEQ ID 4179> which encodes the amino acid sequence <SEQ ID 4180>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3285 (Affirmative) < succ>
10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9785> which encodes amino acid sequence <SEQ ID 9786> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:BAE06554 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 108/211 (51%), Positives = 149/211 (70%)

 Query: 7 EDVILNATERNMVHKLQNDPSGHDMFHVIRVENLAVELAHKEGANTFICQMAALLHDIID 66
 E IL + E V +L ++ SCHDMHI RV +A + +E + F+ Q+AAL HD-ID
20 Sbjct: 3 EQAILQSAEAWVKQLMDEYSCHDMYHIRRVILMAKAIGEQRKVDVFVQIALPFDLID 62

 Query: 67 DKICQDSKQASYELTOMLYSQDLATAEVEHILDILENISFKAGTGLTMKTLEQQIVQDAD 126
 DK+ D + A +L W+ + + ++H +DI+ ISFK G G ++ T E +VQDAD
25 Sbjct: 63 DKLVDDFETAKQQLIDMMSAAGVPSCKIDHIMDIINTISFKGHGQSIATREAMVQDAD 122

 Query: 127 RLDAMGAIGIARTMAYSGSKRLIHDPNLKPRENLLEBYRNQDPTAIHFYKLLKLD 186
 RLD+A+GAIGIART AYS+KG+ I+DP L RE +T+EEYR+G+ TAI HFYKLL KLD
30 Sbjct: 123 RLDALGAIGIARTFAYSNGKQPIYDPELPIRETNTVEEYRHGKSTAINHFYKLFKLD 182

 Query: 187 LMNTIKQEMLAQKRHDFLELYLAEFYAEWNG 217
 LMNT+ GK LA++RH F+E ++ F +EWNG
30 Sbjct: 183 LMNTETGKQLAKZRHVFMBOQFIERFLSEWNG 213

```

No corresponding DNA sequence was identified in *S. pyogenes*.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1368**

A DNA sequence (GBSx1453) was identified in *S. agalactiae* <SEQ ID 4181> which encodes the amino acid sequence <SEQ ID 4182>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 21
 >>> May be a lipoprotein

 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50 !GB:U25448 internalin [Listeria monocytogenes]
 !GB:U25448 internalin [Listeria monocytogenes]
 !GB:U25448 internalin [Listeria monocytogenes]
 !GB:U25448 internalin [Listeria monocytogenes]

 >GP:AAA69530 GB:U25448 internalin [Listeria monocytogenes]
55 Identities = 78/253 (30%), Positives = 132/253 (51%), Gaps = 2/253 (0%)

```

- Query: 531 LKQLMNTNIGITDYSFLDKMPLLEGLDISQNGIKDLSFLTKYKQLSLIAAANNIGTSLKF 590  
L Q+ +N +TD + L + L + ++ N I D++ L L+ + NN IT + P  
Sbjct: 26 LTQINFSNNQITDITPLKDLTKLVDIIMRNQIADITPLANLSNLGLTLPNNQITDIDP 85
- Query: 591 LAELPNLQFLVLSHNNISDLTFLSNLTKLQELYLDDHNNVKNLSALSQKKDLKVLDSNNK 650  
L L NL L LS N ISD++ LS LT LQ+L L N V+L L+ L+ ID+S+NK  
Sbjct: 86 LKNLTNLAKLELSNFTISDISALSGLTSLQQLSLG-NQVTDLKPLANLTFLKRLDISNNK 144
- Query: 651 SADDLSTL-KTTSLETLNINETNISMFLKQNKPKVSNITINNAKLASLDGIEESDRIVKV 709  
+D+S L K T+LE+L+ S+++ L + L++N +L + + + +  
Sbjct: 145 VSDISVLAKLTNLES LIATNNQISDITPLGIIITNLDLSLNGNQLKDGTGTHASLTN+TDL 204
- Query: 710 BAEGNQIKSLVLKKNQKSLKPLFNVTNNQITSLBGVNNYTSLETLSVSKNKLESIDIKTPN 769  
+ NQI +L L L + NQ++++ + T+L L +++N+LE + +  
Sbjct: 205 DLANNQISNLAFLPGLTKLTELKGANQISNIXPLAGLTALTNLELNENQLEDISPI SNL 264
- Query: 770 KTVTNLDFSHNNV 782  
K +T L NN+  
Sbjct: 265 KNLTLYTLYFNNI 277  
Identities = 91/300 (30%), Positives = 141/300 (46%), Gaps = 42/300 (14%)
- Query: 519 INDMTPVLQFKKLKQLMNTNIGITDYSFLDKMPLLEGLDISQNGIKD---LSFLTKYKQL 575  
I D+TP+ L L + N ITD L + L L+S N I D LS IT +QL  
Sbjct: 58 IADITPLANLSNLGLTLPNNQITDIDPLKDLTNLNLKLELSNFTISDISALSGLTSLQQL 117
- Query: 576 SLIAAANNIGTSLKPLA-----ELPNLQFLVLSHNNISDLTFL 613  
SL N +T LKPLA +L NL+ L+ ++N ISD+TFL  
Sbjct: 118 SL----GNQVTDLKPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPL 173
- Query: 614 SNLTKLQELYLDDHNNVKNLSALSQKKDLKVLDSNNKSADLSTLK-TTSELTLLNENIT 672  
LT L EL L N +K++ L+ +L LDL+NN+ ++L+L T L L L  
Sbjct: 174 GILTNDLESLNGNQLKDGTGTHASLTNLTDLANNQISNLAFLPGLTKTELKGANQI 233
- Query: 673 SNLSFLKQNKPKVSNITINNAKLASLDGIEESDRIVKBAEQNKISLVLKNQKQS LKFLN 732  
SN+ L + ++NL +N +L + I + + N I + + L+L  
Sbjct: 234 SNIXPLAGLTALTNLELNENQLEDISPI SNLKYTLYTLYFNNISDISPVSSLTKLQRLP 293
- Query: 733 VTNNQITSLBGVNNYTSLETLSVSKNKLESIDIKTPNKTVTNLDPSHNNVPTSQLKNEK 792  
NN+++ + +N T++ LS N++ L TP +T + +QL LN++  
Sbjct: 294 PYNNKVSUVSSLAHLATNINWLGAGHNOISDL---TPLANLTRI-----TOLGINDQ 341  
Identities = 73/253 (28%), Positives = 124/253 (48%), Gaps = 4/253 (1%)
- Query: 540 GITDYSFLDKMPLLEGLDISQNGIKDLSFLTKYKQLSLIAAANNIGTSLKPLAELPNLQF 599  
GI L+ + L + +S N +D++ L +L I NN I + P L A I NL  
Sbjct: 13 GIKSIDGLEVIANLTQINFSNNQITDITPLKDLTKLVDIIMRNQIADITPLANLSNLG 72
- Query: 600 LVLSHNNISDLTFLSNLTKLQELYLDDHNNVKNLSALSQKKDLKVLDSNNKSADLSTLK 659  
L L +N I+D+ PL NLT L L L N + ++SALGG L+ L LN +  
Sbjct: 73 LTLFNQITDIDPLKDLTNLNLKLELSNFTISDISALSGLTSLQQLSLGNGVTDLKPLAN 132
- Query: 660 TSLETLIANNTNINLSPLKQNKPKVSNITINNAKLASLDGIEESDRIVKBAEQNKISL 719  
T+LE L ++ S++S L + +L N +++ + + ++ GNQ+K +  
Sbjct: 133 TTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGIIITNLDLSLNGNQLKDI 192
- Query: 720 VLNNKQKSLKPLFNVTNNQITSLBGVNNYTSLETLSVSKNKLESIDIKTPNKTVTNLDPSH 779  
+L L++ NNQ++L + T L L + N++ ++ +TNL+ +  
Sbjct: 193 GTLASLTNLTDLANNQISNLAFLPGLTKTELKGANQISNIXPLAGLTALTNLELN 252
- Query: 780 NNV---PTSQLK 788  
N + P S LK  
Sbjct: 253 NQLRDISPISNLK 265  
Identities = 56/209 (26%), Positives = 115/209 (54%), Gaps = 2/209 (0%)
- Query: 575 LSLIAAANNIGTSLKPLAKLAPLQFLVLSHNNISDLTFLSNLTKLQELYLDDHNNVKNLS 634  
++ + A GI S+ L L NL + S+N ++D+TFL +LTKL ++ ++N + ++  
Sbjct: 4 VTTLQADRLGKSIDGLEVIANLTQINFSNNQITDITPLKDLTKLVDIIMRNQIADITP 63

-1495-

Query: 635 LSGKKDLKVLDSNNKASADLSIACT-TSLETLINNETWNTSNLSPKQNPVSNLTINNAK 693  
 L+ +L L L NN+ D+ LK T+L L L+ S++S L + L++ N +  
 Sbjct: 64 LANLSNLGLTLPNNQITDIDPLKNLTNINRLSSNTISDISALSG/LTSLQQLSLGN-Q 122

5 Query: 694 LASLDGIEESDEIVKVEBQNKISVLNKKQGSLLKPIAVTNHQLTSLEGVNNYTSLETL 753  
 + L + + + + + N++ + + K +L+ L TNQ++ + + T+L+ L  
 Sbjct: 123 VTDLPLANLTLERLIDISSNKVSDISVLAKLTNLESLLDATTNNQISDITPLIGILTDLDEL 182

10 Query: 754 SVSNKLLSSLDIKTPANKITVINLDPSSNNV 782  
 S++ N+L+ + +T+LD ++N +  
 Sbjct: 183 SLNNGQLKDIGTSLASLTNLTDLQANQI 211  
 Identities = 61/228 (26%), Positives = 118/228 (51%), Gaps = 3/228 (1%)

15 Query: 483 LATVITKINIGQRTNPFQRFGLMPEINIEVLIGFTPIINDMTFVLQFKKLQWLMTNIGIT 542  
 L+ + + + + G + + L+ + +E L I + +D+ + + L+ L TN I+  
 Sbjct: 111 LLSLQQLSLGNQVTDLKP--LANLTLERLIDISSNKVSDISVLAKLTNLESLLDATTNNQIS 168

20 Query: 543 DYSFLDKMPLLEGLDISQNGIKDSFLTYYKQLSLIAAANNQITSLKPLAELPNLQFLVL 602  
 D + L + L+ L + + N +RD+ L L+ + ANN I+L PL L L L L  
 Sbjct: 169 DITPLGILTINLDELSSANGQLKDIGTSLASLTNLTDLQANQISMLAPLPGITKLTELK 228

Query: 603 SHRNISDLTPLSNLTQLQELYLDHNNVKNLSALSGKKDLKVLDSNNKASADLSIACT-TS 661  
 N IS++ PL+ LT L L L+ N + + + + S +S K+L L L N +D+S + + T  
 Sbjct: 229 GANQISNIXPLAGLTALTNLELNENQLEDISPISNLANLTLYTLFPNNISDISPVSSLT 288

25 Query: 662 LETLLINNETWNTSNLSPKQNPVSNLTINNAKLSLDGIEESDEIVK 709  
 L+ L S++S L + +L+ + + + L + I + +  
 Sbjct: 289 LQRLFFFTNNKVDVSSLANLTNINWLSAGHNSIDLTPLANLTRITQL 336  
 Identities = 60/286 (20%), Positives = 129/286 (44%), Gaps = 24/286 (8%)

30 Query: 369 SNKLSDEDQKKLTYLAELKGINPQIEVLTSEDGSIIFKYPHDDHSHITLAKDIEIGKPI 428  
 +N++D D K + +L L+ N I + + + G + + + +G +  
 Sbjct: 77 NNQITDIDPLKNLTNINRLSSNTISDISALSG-----LTSLLQQLSLGNQV 123

35 Query: 429 PDGHHDSHAKKVGMAFLKQIGFDEIIGDILHADAPTFFPNNINPEKMRG--LATV 486  
 D K + TL++ + DI T S + + L+  
 Sbjct: 124 TD-----LKPLANLTLERLIDISSNKVSDISVLAKLTNLESLLDATTNNQISDITPLGIL 176

40 Query: 487 TKIN-IGQRTNPFQRFGLMPEINIEVLIGFTPIINDMTFVLQFKKLQWLMTNIGITDY 544  
 T + + + N + G L+ + N+ L+ L+ I++ P+ KL + L + I++  
 Sbjct: 177 TNLEDELSSANGQLKDIGTSLASLTNLTDLQANQISMLAPLPGITKLTELKLANQISNI 236

Query: 545 SFLEKMPLEGLDISQNGIKDSFLTYYKQLSLIAAANNQITSLKPLAELPNLQFLVL 604  
 L + L L + + + N +D+S + + K L+ + N I+ + P+ L LQ L L  
 Sbjct: 237 XPLAGLTALTNLELNENQLEDISPISNLANLTLYTLFPNNISDISPVSSLT/KLQRLFFYN 296

45 Query: 605 NNISDLTPLSNLTQLQELYLDHNNVKNLSALSGKKDLKVLDSNNK 650  
 N +SD+ L+NLT + L HN + +L+ L+ + L L+ + +  
 Sbjct: 297 NKVSDVSSLANLTNINWLSAGHNSIDLTPLANLTRITQLGLANDQ 342

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4183> which encodes the amino acid sequence <SEQ ID 4184>. Analysis of this protein sequence reveals the following:

Possible site: 21

55 >>> May be a lipoprotein

----- Final Results -----

60 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA69530 GB:U25448 internalin [*Listeria monocytogenes*]  
 Identities = 88/279 (31%), Positives = 149/279 (52%), Gaps = 2/279 (0%)

65

-1496-

|                                                             |                                                                             |
|-------------------------------------------------------------|-----------------------------------------------------------------------------|
| Query:                                                      | 419 L PNLSTLIGIGFTPIKDISFVLQFKKLQLLMTKTGVTDYRFLDNMPQLBGLDISQNNLKD 478       |
| L + TL                                                      | IK I + L Q + + +TD L ++ +L I ++ N + D                                       |
| Sbjct:                                                      | 1 LDXVTTLQADRIGIKSIDGLSEYLNINLTQINFNNQLTDTTFKDLTKVDILMNNQIND 60             |
| 5                                                           | Query: 479 ISPLSKYKNTLVAADNGIEDIRPLQGLPNLKFVLNNKISDLASPLASLQQLQLHID 538     |
| I+ L+ NLT + +N I DI EL L NL L LS+N ISD+S L+ L IQ+L +        |                                                                             |
| Sbjct:                                                      | 61 ITPANLNLNLTGLTFNNQITTDIDPLKNTLMARLELSNTISDISALSGSLTSIQQGLL- 119          |
| 10                                                          | Query: 539 NQITDLSFVSHKBSLTVDLSRNADVLDLTL-QAPKLETLMVNDTKVSHLDFLKNNNL 597    |
| NQ+TL P+++ +L +D+S N D++ L+ + LS+L+ + +S + L NL             |                                                                             |
| Sbjct:                                                      | 120 GNVDTLKLPLNLTTLRELDSISGNKVDISVLAKI/TNLSL IATNQISDTPLGLITL 179           |
| Query:                                                      | 598 SSLSINRAQLQSLBGTASSVIVRVEABGNQIKSLVLKDKQSGSLTFLDVTGNQLTSLSCV 657        |
| LS+N QL+ + + + + ++ NQI +L LT L+ + MQ++++ +                 |                                                                             |
| Sbjct:                                                      | 180 DELSLNGNQLKDIQTGLASLTNLTDLDLANNQIENLAPLGLTKLTELKLGANQISLXPL 239         |
| Query:                                                      | 658 NNPTALDILSVSRNQITLVNLSKPKNTVTYNIDISERNIS 696                            |
| TAL L +++NQL +++ K +T + + NRTS                              |                                                                             |
| Sbjct:                                                      | 240 AGTALTNLNLNLRNQLDISPTNLSNLTUTLTYPNTIS 278                               |
| 20                                                          | An alignment of the GAS and GBS proteins is shown below.                    |
|                                                             | Identities = 346/753 (46%), Positives = 472/753 (61%), Gaps = 63/753 (8%)   |
| Query:                                                      | 187 SRLGNQSNHYRVNSK-----IAGLHYPTNGFLFNGRG-IKGTTPTGILVSHNN 237               |
| SR G SN + SK +AG+ +PT +GF+ I T GI+V+H H                     |                                                                             |
| Sbjct:                                                      | 38 SRGKMTSNKIKPIKSKCKTNKTHKGVAGVDFPTDDGFILTKDKLSKTDGGIVVDHGH 97             |
| Query:                                                      | 238 LHFISFADLRKGGW-----GSIADRYQPKKADSKKQPSKPKPTENTLPKDI--NDK 289            |
| HFI +ADL+ + G+ + ++A S+ S + F DI +D                         |                                                                             |
| Sbjct:                                                      | 98 SHFITVADLKGSPFEYLI PKGASLAKPAVQRAASQGTSKVADPHHHYEFNFADIVAEDA 157         |
| Query:                                                      | 290 LAYLARE---LHLDI-----SRIRVLATLNGEIGPFTADDHT 324                          |
| L Y R H + S + TNG G +P D                                    |                                                                             |
| Sbjct:                                                      | 158 LGYTVRHDDHPHYLIKSSLGCGTQAQAKQVATRLPQSSSLVSTANGISGLHFFPTSDGF 217         |
| 35                                                          | Query: 325 HVIMAKDIDLKFPIMPHDDEDH-----HKGHHHD--ESDHKESEHETK 368             |
| + ++ HD H H +D +++ E H+ +                                   |                                                                             |
| Sbjct:                                                      | 218 QFNQGIIVGVTKDSILVNHGDHLHPISFADLRQGGWARVADQYDFAKAEKPASTHQTP 277          |
| 40                                                          | Query: 369 SNKLSRSDQKKLIYLASKLGLNPNQIEVLTSRSDGSIIPKYPHHDSHTIASDKIEIGHPI 428 |
| + + E Q+KL YLASKLG+P+ I+ + ++DG + +YFH D+H + DSDZG I        |                                                                             |
| Sbjct:                                                      | 278 LSEREKETQKXLYLASKLGIDPSTIKRVETQDGKLGLEYPHHDAHVLMLSPITEIGDI 337          |
| 45                                                          | Query: 429 PDGH---HDHSHAKDKVGMATLQKIGDPDRI IQDILHA-DAPTPPSPNETNPEHQMQLA 484 |
| PD H H K KVM TL+ +QPD+E+I DI- DAPTPPSPNE +P MA+HLA          |                                                                             |
| Sbjct:                                                      | 338 PDPAHAEHARELEKHKVMDTLRALGPDSEVILDIRTHDAPTPPSPNEKDPNMWKEWLA 397          |
| Query:                                                      | 485 TVTKINIQORTNPPORFGLSLMNPTEIVLQIGFTPIINDMTPVLQFKKLQOLAMTNGITDY 544       |
| TV K++G R +P OR GLSL+PM+E LGIGFTPI D++PVLQFKKLKOL MT TG+TDY |                                                                             |
| Sbjct:                                                      | 398 TVIKLIDLSGRKDPQRKGLSLPMLSTIGIGFTPIKDISFVLQFKKLQLLMTKTGVTDY 457          |
| 50                                                          | Query: 545 SFIDMPLBGLBGLDISQNGIKDLSPVLYKYLQSLIAAANGITSLKPLARLPLNQLFVLSH 604 |
| FLD MP LBG+DISQN +D+SFL+KYL L-L+AAA+NGI ++PL +LNL+FLVLG+    |                                                                             |
| Sbjct:                                                      | 458 RFLDNMPQLBGLDISQNNLKUISPLSKYKNTLVAADNGIEDIRPLQGLPNLKFVLSH 517           |
| 55                                                          | Query: 605 NNISDLTFLSNLTKLQSLYLDHNNVKNLSALSGKDKLVLDLNNKNSADLTATTSLEP 664    |
| N ISDL+PL++L +LQEL+D+N + +LS +S K+ L+V+DLS N DL+TL+ LET     |                                                                             |
| Sbjct:                                                      | 518 NKISDLSPASLHQLQLHIDNNQITDLSFVSHKBSLTVDLSRNADVLDLTLQAPKLET 577           |
| 60                                                          | Query: 665 LLNETNTNLSLFLKQNKPKVSNLTINNAKLASLUGIEESDIIVRVEABGNQIKSLVLNKK 724 |
| L++N+T S+L FLK NP +S+L+IN A+L SL+GIE S IV+VABGNQIKSLVLN+K   |                                                                             |
| Sbjct:                                                      | 578 LAMNNTVKSHLDFLKNNNLSSLINRAQLQSLBGTASSVIVRVEABGNQIKSLVLNKK 637           |
| Query:                                                      | 725 QGSLKPLNVTNMQITSLGVNNYTSLETLSVSKNKLSDIKTPNKTPTNLDPSHNNVPT 784           |
| QGGL PL+VT NQLTSLEGVNN+T+L+ LSVSN+L ++++ PKRTYND+ SHNN+     |                                                                             |
| Sbjct:                                                      | 638 QGSLTFLDVTNMQITSLGVNNPTALDILSVSRNQITLVNLSKPKNTVTYNIDISERNISL 697        |

-1497-

Query: 785 SQLKLNKNIPEAVAKNFPVAVGSMVGVGSLAEKQAMASKKDKQVSD -NTNHQKNTERK 843  
 + LKLNK++IPEA+AKNFPVAVGSMVGVG+ EKQAMA+K + ++H N +  
 Sbjct: 698 ADLKLNKCHIPEATAKNFPVAVGSMVGVGTASEKAAATKAKESAQRAESHEDYNNHT 757

5 Query: 844 AQAQVADSKKENPKTHDEHHDEESTDHAHVGVHH 876  
 + E+ D H+HE+ +A +H  
 Sbjct: 758 YEDESCHAHHRDKDDHDEHREHREKAEKDEQNH 790

SEQ ID 4182 (GBS84) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 9; MW 97.6kDa).

GBS84-His was purified as shown in Figure 194, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1369

15 A DNA sequence (GBSx1454) was identified in *S. agalactiae* <SEQ ID 4185> which encodes the amino acid sequence <SEQ ID 4186>. This protein is predicted to be GTP-binding protein lepa (lepA). Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1962 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE14493 GB:Z99117 GTP-binding protein [Bacillus subtilis]  
 Identities = 464/603 (75%), Positives = 540/603 (88%)

30 Query: 8 KROEKIRNFSIIAHIDHGKSTIADRILEKTETVSRSKQQLLSDMDLERERGITTIKNA 67  
 +RQ +IRNFSIIAHIDHGKSTIADRILEKT ++ RRM+ QLLSDMDLERERGITTIKNA  
 Sbjct: 9 ERQSRIRNFSIIAHIDHGKSTIADRILEKTSAITOREMKQQLLSDMDLERERGITTIKNS 68

35 Query: 68 IEIANTXNDGETYIFHLIDTPGHVDFTYEVSRSLAACGAILVDDAQGISACTLANVYL 127  
 ++L Y AKDGE YIFHLIDTPGHVDFTYEVSRSLAACGAILVDDAQGISACTLANVYL  
 Sbjct: 69 VOLKXKAKDGEYIFHLIDTPGHVDFTYEVSRSLAACGAILVDDAQGISACTLANVYL 128

40 Query: 128 ALNDLLEILPVINKIDLPAADPERKVRARVEDVIGLDASAVLASAKAGIGISEILEQIVE 187  
 ALNDLLEILPVINKIDLPA+PERVR EVEDVIGLDASAVLASAKAGIGISEILEQIVE  
 Sbjct: 129 ALNDLLEILPVINKIDLPASAEPRVRQVEDVIGLDASAVLASAKAGIGISEILEQIVE 188

45 Query: 188 KVAFTQTEVDPAQLQALIFDGVYDAYRGVILLQVRIVKMWKPGDKIQMNSNGKTDPTVEVG 247  
 KVAFTG+ +APL+ALIFDS+YDAYRGV+ +R+V G VKGG KI+MM+ GK F+VTEVG  
 Sbjct: 189 KVAFTQTEKPAFLKALIFDSIDYATRGVAYIRVVEGTVKRGQIKMWATGKEFEETVEG 248

50 Query: 248 IFIPRAWGRDPLATGSDVGYIASGKIVADTRVGDITLANNPAIDELHGYKQKQNPVAVG 307  
 +FTRPA + L GDWG+ ASIK V DTRVGDIT A NPA E L GY+ ++NPNV+ G  
 Sbjct: 249 VFTPKATFTNELTVGDVGFLTASTIKVNGVTRKVGDTTSSAANPAEELPGRKLNTPVTCG 308

55 Query: 308 LYPIESKYNLDLRALEKLELNDASLQPEKTSQALGPGFCGFLGLLHMVIOERLEIE 357  
 LYPI++ KYNLDLRALEKLELNDASLQPEKTSQALGPGFCGFLGLLHMVIOERLEIE  
 Sbjct: 309 LYPIDATKYNLDLRALEKLELNDASLQPEKTSQALGPGFCGFLGLLHMVIOERLEIE 358

Query: 368 FNIDLIHTAPSVYVHVNTDGEMLVENSNESEFPDPTRVDSIEEPVVAQIMVQEVAV 427  
 FNIDLI TAPSV+Y V TDGE + V NES FDP +++ +EFPVVA +NP +VQAV  
 Sbjct: 369 FNIDLIHTAPSVYVHVNTDGEKVVVNDSENNPDPQKIBERVEEFTVVAQIMVQEVAV 428

Query: 428 MELAQKRGDFVTMDYIDNRVHVYIQIPIARIVDFDFDKIKSSTRGYASDFYSEISYR 487

-1498-

MEL Q KRG+F+ M Y+D NR++IY +FLAEIV++FPD+LKST+GYASFQY+ Y+  
 5 Sbjet: 429 MELQXGKGRNFIDMQYLDANRVSIYDMFLAEIVYBFPDQLKSTKGYASFQYRLIGYKP 488

Query: 488 SQLXKMDILLANGDKVDALSFVHKEFAYERGKLIYDKLKIITPRQQFVPTQAALGQKIV 547  
 S+L XMDI+LNG+K+DALSFVH++AYERK+IV+KLK++IPRQQFVPTQAALGQKIV  
 10 Sbjet: 489 SKIWKMDIMLNGEKIDALSFVHRDYAYERGKVIYKRLKRLIPRQQFVPTQAALGQKIV 548

Query: 548 ARSDIKALRKNVLAQCYGSDVSRKKRILEKQKAGKRMKAIGSEVFPQEAFLSVLMDDD 607  
 ARS IKA+RKNVLAQCYGSD+SRKKRILEKQK GK+RMK +GSVEVTPQEP++VL MDD  
 15 Sbjet: 549 ARSTIKAMRKNVLAQCYGSDISRKKRILEKQKQKGRKMKQVGSVEVTPQEPMAVLKMDSD 608

Query: 608 DKK 610  
 KK  
 Sbjet: 609 PKK 611

A related GBS sequence was identified <SEQ ID 10775> which encodes the amino acid sequence <SEQ ID 10776>. A further related GBS nucleic acid sequence <SEQ ID 10955> which encodes amino acid sequence <SEQ ID 10956> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4187> which encodes the amino acid sequence <SEQ ID 4188>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1829 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GF: CAB14493 GB: Z99117 GTP-binding protein [Bacillus subtilis]  
 Identities = 463/603 (76%), Positives = 542/603 (89%)

Query: 8 KRQEKIRNFSTIAIHIDHGKSTLADRILEKTETVSSREMQQLLDSMDLERERGITIKLNA 67  
 +RO +IRNFSTIAIHIDHGKSTLADRILEKT ++ REM+ QLLDSMDLERERGITIKLN+  
 35 Sbjet: 9 ERQSRIRNFSTIAIHIDHGKSTLADRILEKTSITQREMKBQLLDSMDLERERGITIKLNS 68

Query: 68 IELNNTAKDGETYIFHLIDTPGHVDFTYEVSLSLAACEGAILVVDAAQGIQAQTLANVYL 127  
 ++L Y AKDGE YIFHLIDTPGHVDFTYEVSLSLAACEGAILVVDAAQGIQAQTLANVYL  
 40 Sbjet: 69 VQLKKYKADGEBEYIFHLIDTPGHVDFTYEVSLSLAACEGAILVVDAAQGIQAQTLANVYL 128

Query: 128 ALNDNLEILPVINKIDLPADPERVREVEDVIGLDAEAVLASAKAGIGIEILBEQIVE 187  
 ALNDNLEILPVINKIDLP+ +PERVR EVEDVIGLDAEAVLASAKAGIGIEILBEQIVE  
 45 Sbjet: 129 ALNDNLEILPVINKIDLPASRPERVQEVEDVIGLDAEAVLASAKAGIGIEILBEQIVE 188

Query: 188 KVPAPTGDVDAFLQALIFDSVYDAYRGVILQVRVINGIVKPGDKIQMNSNGKITFDVTEVG 247  
 KVPAPTGD +APL+ALIFDS+YDAYRGV+ +R+V G VKPG KI+MM+ GK F+VTEVG  
 Sbjet: 189 KVPAPTGDPEAPLAKALIFDSINAYRGVAVYRVVBGTGKPGQKIKMMATGKBEVTEVG 248

Query: 248 IFTPKAVGRDFLATGDVGYVAASIKTVADTRVGDVTTLNANFAKALHGYQNMNKFVAG 307  
 +FTPKA + L GDVG++ ASIK V DTRVGD+T A NFA+KAL GY++NMN+ G  
 50 Sbjet: 249 VFTPKATVYNISLVGDVGFLTASIKNVGDTVRVGDITTSANFAEALPGYKINMNVYAG 308

Query: 308 IYPIESNKYNDLREALEKILNDASLQFEPETSQALGFGPRGFLGLLHMDVIOERLRE 367  
 +YPI+ KYNDLREALEKILND+SLQ+E ETSQALGFGPRGFLG+IHM++IQER+ERE  
 55 Sbjet: 309 LYPIDTAKYNDLREALEKILNDASLQYBAETSQALGFGPRGFLGLHMDVIOERLRE 368

Query: 368 FNIDLINTAPSVYHVHETTDIMIEVSNPSEFPDPTVRVAFIEEPYKQIMVPQEFVGAV 427  
 FNIDLI TAPSV+Y V+ TD + + V NPS PDP ++ +ERPVKA +MVP ++VGAV  
 60 Sbjet: 369 FNIDLITAPSVIYDVMYTDGKVVVDNFSNMPDQKIERVEKPEYKATVNPVNDYVGAV 428

Query: 428 MELSQKRGDFVIMYIDIDNRVNVIVYFLAEIVDFPDQLKSTKGYASFQYDYNSEYR 487